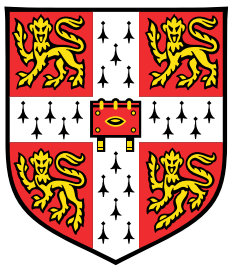
On Latent Variable Models for Bayesian Inference with Stable Distributions and Processes



Marina Riabiz

Department of Engineering
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Clare Hall College October 2018

To my family

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Marina Riabiz
October 2018

Acknowledgements

At the end of this work, I feel obliged to acknowledge, or at least try to, all the people that let me reach this point.

I would first like to thank my Supervisor, Prof. Simon Godsill, for accepting me on a PhD course in Cambridge, and for guiding and encouraging me throughout my research on such an interesting topic. He has been able to make me focus on the big picture while providing invaluable advice on the details, allowing me to keep an overall balance, and move forward when I felt stuck. None of this thesis would have been possible without him.

I would then like to thank the collaborators I had during this journey. First of all, Dr. Tatjana Lemke, that had a large patience in putting me on the right track at the beginning of my PhD on stable distributions. I would then like to thank Dr. Fredrik Lindsten, for guiding me during my first year of PhD, and for generously sharing his time whenever I had a question: I learned a lot from that. The second part of my PhD was influenced by the analytical approach of Dr. Tohid Ardeshiri, who thought me how to face problems in a more systematic way, spending long hours discussing and writing maths: I am grateful to him tor this but also for his friendship. Finally, the closed-form bounds established in this thesis would not be available without the contribution of Prof. Ioannis Kontoyiannis. I would like to thank him for taking interest in the topic, working on it, and for co-supervising me in the last year, by being available for any discussion and question. I had a glimpse into what sharp, mathematical thinking looks like.

I am grateful to the first two (official) readers of this thesis, my examiners, Prof. Franccois Septier and Prof. Joan Lasenby for having the patience to go through this work and providing useful insights on it.

I would like to thank the people that made my life good and unique in its way in Cambridge. First of all, I would like to thank Dr. (and Kuma) Milica Gasic, because

she is part of the reason I thought about applying to Cambridge, and because she has been a guide for me during these years. Her advice and support (from a shoulder any time I needed to a countless number of Ottolenghi dinners) are part of the engine that pushed me through this long way. I would also like to thank all the present and past members of the SigProC group (and other groups on our floor, and, more generally, in the Engineering Department) for creating a nice working atmosphere that I will miss. Moreover I would like to thank the kind but professional administrative support that I have had in my PhD: thank you Rachel, Louise, Diane, Laura and Rob. I would then like to thank my housemates throughout the years and my Clare Hall College friends, in particular the College Social group, because they have been an invaluable alternative to the Department, any time a working session with tea (and/or bacon) was needed. Being part of the Yugosphere Society in Cambridge has changed my perception of my identity, by making me re-discover a part of my life that I was missing before coming here, so I would like to thank all the happy friends that I have met along the way. I would also like to thank all the people that I met through various activities, from rowing, to singing, to attending yoga classes, as well as the students that I thought, and from whom I learned a lot. I am also grateful for the possibility I was given to spend an 'enrichment year' at the Alan Turing Institute, for the friends I made, and the (future) collaborators that I met. Finally I would like thank this journey for the allowing me to travel and meet wonderful people at a number of conferences and workshops.

My most sincere thanks goes to my family, for the constant support throughout my life. What I wrote in my Master thesis holds, and I would like to repeat it here, because it is like a fixed point for me. Thanks to Dejan, because he constantly teaches me that life needs to be taken less seriously, and that it is always worth spending oneself for others. Thanks to Ines, because she is a model for the firmness and determination that characterize her, and make her special. Thanks to our parents: acrobatic jumps are nothing, compared to the efforts you have made and which you would still be willing to do for us. Thanks to my mom, because she teaches me that I need to stand behind my choices and because she the glue of our family. Thanks to my dad because of his natural curiosity towards everything: he is one of the few people with whom I can talk about maths and jump to literature two minutes later. Every step forward I do is also yours.

Finally I would like to thank Nikolay, for being with me and there for me in the last year and a half, with patience and constant support. The end of my PhD would have been much harder without him, but now that this has almost come to an end, I am looking forward to what the future holds for us.

Abstract

Extreme values and skewness in time-series are often observed in engineering, financial and biological applications. This thesis is a study motivated by the need of efficient and reliable Bayesian inference methods when the α-stable model is selected to represent such data.

The class of stable distributions is the limit of the generalized central limit theorem (CLT), having a key role in representing phenomena that can be thought of as the sum of many perturbations, with potentially unbounded variance. Besides the ability to model heavy-tailedness, another consequence of the generalized CLT is a further degree of freedom of stable distributions, namely their potential skewness. However, stable distributions are, at the same time, highly intractable for inference purposes. Several approximate methods are available in the literature, in both the frequentist and Bayesian paradigms, but they suffer from a number of deficiencies, the greatest of which is the lack of quantification of the approximation in place. This thesis proposes Bayesian inference schemes for two different latent variable models, with the aim of providing guarantees of accuracy when the α-stable model is used.

In the first part of the thesis, a marginal representation of the α-stable density is used to develop a novel, asymptotically exact, Bayesian method for parameter inference. This is based on the pseudo-marginal Markov chain Monte Carlo (MCMC) approach, that requires only unbiased estimates of the intractable likelihood, computed through adaptive importance sampling for the marginal representation. The results obtained are comparable to a state of the art conditional Gibbs sampler, but do not introduce any approximation, while allowing for better control of the quality of the inference.

The focus of the second and central part of the thesis is the Poisson series representation (PSR) of α-stable random variables. An approach that turns the infinite-dimensional PSR into an approximately conditionally Gaussian representation, by means of Gaussian approximation of the residual of the series, has been presented in previous literature, together with inference procedures such as MCMC and Particle Filtering. In this setting, the first contribution of this dissertation is the formulation of a CLT for the PSR residual, which serves to justify the existing approximation.

Moreover, numerical and theoretical results on the rate of convergence for finite values of the truncation parameter are presented. The convergence is examined directly in terms of Kolmogorov distance between distribution functions, through the application of probability theoretic results, such as the Esséen's smoothing lemma. This analysis allows for the selection of appropriate truncations for different α-stable parameter configurations and gives theoretical guarantees on the accuracy achieved when using the PSR model. Furthermore, superior behaviour of the proposed approximation is found, compared to the simple series truncation, justifying its use for inference tasks.

In the third and final part of this thesis, an extension of the modified Poisson series representation (MPSR) of linear continuous-time models driven by α-stable Lévy processes to the multivariate case is presented. Stable Lévy processes are suitable to model jumps and discontinuities in the state, while possessing the self-similarity property, which makes these processes a very natural class for the driving noise in continuous time models. A scheme for approximate simulation from the multivariate linear models, namely multivariate stable vectors evolving in time, is presented. While stable random vectors are parametrized by a function, the presented approximate approach involves only finite dimensional parameters. This will facilitate inference methods, to be developed in future work, towards which the proposed simulation methods constitute the foundational work.

Contents

List of Figures

List of Tables

List of Algorithms

Notation

Greek Symbols

α	Tail parameter of the stable distribution
σ	Scale parameter of the stable distribution
β	Skewness parameter of the stable distribution
μ	Location parameter of the stable distribution
$\boldsymbol{\theta}$, $\boldsymbol{\lambda}$	Vectors of parameters
Θ	Parameter space

Other Symbols

$X \sim \mathcal{D}$	The random variable X has distribution $\mathcal{D}$
$X \stackrel{\mathcal{D}}{=} Y$	The random variable X has representation Y
$X \stackrel{\mathcal{D}}{\to} \mathcal{D}$	The random variable X converges in distribution to $\mathcal{D}$
$X \overset{\text{approx}}{\sim} \mathcal{D}$	The distribution of the random variable X is approximated by $\mathcal{D}$
$\mathcal{S}_\alpha(\sigma, \beta, \mu)$	Stable distribution with parameters α, σ, β, μ
$\mathcal{S}_{\alpha,d}(\boldsymbol{\mu}, \xi)$	Multivariate (d-dimensional) stable distribution with tail parameter α, location $\boldsymbol{\mu}$ and spectral measure ξ
$\mathcal{N}(\mu, \sigma^2)$	Gaussian distribution with mean μ and variance σ^2
$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	Multivariate Gaussian distribution with mean vector μ and variance-covariance matrix Σ
$\mathcal{U}(a, b)$	Uniform distribution on (a, b)
$\Gamma(\alpha, \beta)$	Gamma distribution with shape α and rate β
$\mathcal{E}(\lambda)$	Exponential distribution with mean $1/\lambda$
Lévy(μ, σ)	Lévy distribution with location μ and scale σ
Poisson(t)	Poisson distribution with mean t
IG(α, β)	Inverse gamma distribution with shape α and scale β
$\phi_X(s)$	Characteristic function of the random variable X evaluated in $s \in \mathbb{R}$
$f(x) \approx g(x)$	The function $f(x)$ is approximated by $g(x)$
$F_X(x)$	Cumulative distribution function of the random variable X evaluated in $x \in \mathbb{R}$

$p_X(x)$	Probability density function of the random variable X evaluated in $x \in \mathbb{R}$
$p(\mathbf{x} \mid \boldsymbol{\theta})$	Likelihood function
$\hat{p}(\mathbf{x} \mid \boldsymbol{\theta})$	Monte Carlo estimator of the likelihood function
$\hat{\ell}(\mathbf{x} \mid \boldsymbol{\theta})$	Logarithm of the Monte Carlo estimator likelihood function
$p(\boldsymbol{\theta})$	Prior PDF of $\boldsymbol{\theta}$
$p(\boldsymbol{\theta} \mid \mathbf{x})$	Posterior PDF of $\boldsymbol{\theta}$
$\mathbb{E}[\cdot]$	Expectation
$\mathrm{Var}[\cdot]$	Variance or variance-covariance matrix
$\mathrm{Corr}(\cdot, \cdot)$	Correlation
$\mathrm{Cov}(\cdot, \cdot)$	Covariance
$\mathbb{P}(\cdot)$	Probability
$\Delta(S, T)$	Kolmogorov distance between the distributions of the random variables S and T
$\delta(S, T)$	Distance between the PDFs of the random variables S and T
$i(S, T)$	Uniform bound on the distance between the PDFs of the random variables S and T
$q(S, T)$	Numerical simulations of $i(S, T)$
$I(S, T)$	Smoothed bound on Kolmogorov distance between the distributions of the random variables S and T
$Q(S, T)$	Numerical simulations of $I(S, T)$
$\bar{I}(S, T)$	Bound on Kolmogorov distance between the distributions of the random variables S and T
$\bar{Q}(S, T)$	Numerical simulations of $\bar{I}(S, T)$
$B_i(p_1, p_2, \ldots)$	i-th theoretical bound on a Kolmogorov distance, depending on the parameters $p_1, p_2, \ldots$
UB_{max}	1, the maximum Kolmogorov distance between distributions
$\mathrm{ISE}(f_S, f_T)$	Integrated squared error between the probability density functions of the random variables S and T
$f_1(x) \sim f_2(x)$	The (deterministic and real) function f_1 is asymptotically equivalent to f_2, for $x \to x_0$. If not stated otherwise, $x_0 = \infty$
$f_1(x) = O(f_2(x))$	$\mid f_1 \mid$ is asymptotically bounded above by f_2, for $x \to x_0$. If not stated otherwise, $x_0 = \infty$
$f_1(x) = o(f_2(x))$	f_1 is asymptotically dominated by f_2, for $x \to x_0$. If not stated otherwise, $x_0 = \infty$
i	Unit imaginary number $\sqrt{-1}$
$\mathbb{1}(\cdot)$	Indicator function of the event $(\cdot)$
$\gamma(s, x)$	Lower incomplete gamma function

$\Gamma(s, x)$	Upper incomplete gamma function
$\Gamma(s)$	Gamma function
$\mathbb{N}$	The set of natural numbers
$\mathbb{R}$	The set of real numbers
$\mathbb{C}$	The set of complex numbers
$\{\ell(t)\}$	Non-Gaussian Lévy process
$\text{sgn}(x)$	Sign function
$\delta_a(x)$	Dirac delta function centred at a
G	Number of points adaptively (non i.i.d.) sampled to obtain the LV proposal in the PM sampler
M	Number of points i.i.d. sampled from the LV proposal, to obtain the likelihood estimate in the PM sampler
L	Number of evaluations of LV full conditional distribution used to obtain the likelihood estimate in the PM sampler
S_α	Support of prior and proposal distributions for α
S_β	Support of prior and proposal distributions for β
$X_{(0,c)}$	First part of the PSR for stable RVs
$R_{(0,\infty)}$	Residual of the PSR for stable RVs
$\text{vech}(\cdot)$	Half-vectorization of a symmetric matrix

Acronyms / Abbreviations

ACF(k)	Autocorrelation function at lag k
AR(P)	Autoregressive process of order P
ARMA(P, Q)	Autoregressive moving average process of order P and Q
ARS	Adaptive rejection sampling (algorithm)
CAR(P)	Continuous-time autoregressive process of order P
CDF	Cumulative distribution function
CF	Characteristic function
CLT	Central limit theorem
CMS	Chambers Mallows Stuck method for generating stable RVs
DA(α)	Domain of attraction of the α-stable distribution
DNA(α)	Domain of normal attraction of the α-stable distribution
EM	Expectation maximization algorithm
FDD	Finite dimensional distribution
GAA	Gaussian approximation approach
GAMA	Gaussian approximation of moments approach
GS	Gibbs sampler
i.i.d.	Independent identically distributed (random variables)
IACT	Integrated auto-correlation time

iff	if and only if
IS	Importance sampling (scheme)
LV	Latent variable
MAP	Maximum a posteriori
MaSMiN	Mean and scale mixture of normals
MCEM	Monte Carlo expectation maximization (algorithm)
MCMC	Markov chain Monte Carlo
MH	Metropolis-Hastings
MH-GS	Metropolis-Hastings within Gibbs sampler
ML	Maximum likelihood
MPSR	Modified Poisson series representation
OU	Ornstein-Uhlenbeck
PDF	Probability density function
PM	Pseudo-marginal
PM-MCMC	Pseudo-marginal Markov chain Monte Carlo
PM-MH	Pseudo-marginal Metropolis-Hastings
PM-MH-GS	Pseudo-marginal Metropolis-Hastings within Gibbs sampler
PSR	Poisson series representation
RM	Random measure
RS	Rejection sampling (algorithm)
RV	Random variable
SDE	Stochastic differential equation
SI	Stochastic integral
SLLN	Strong law of large numbers
SMiN	Scale mixture of normals
SSM	State-space model
VARMA(P, Q)	Vector-valued autoregressive moving average process of order P and Q

Chapter 1

Introduction

1.1 Modelling with Heavy Tails

STATISTICAL modelling and inference for random processes is one of the most important tasks in many application areas of science and engineering. In such applications, the evolution over time (or space) of relevant quantities is typically represented through a regression model, in state-space or other forms, which include random noise components, both through the randomness inherent in the underlying system and through the noise in the observation process and measuring instruments.

In order to motivate the theoretical analysis and methods developed in this dissertation, we first consider a simple linear regression model.[1] The linear regression model for a time series $\mathbf{x} := [x_1, \ldots, x_N]'$, describing a particular system realization, can be expressed as

$$\mathbf{x} = \mathbf{G}\boldsymbol{\lambda} + \mathbf{u}, \tag{1.1}$$

where $\boldsymbol{\lambda} := [\lambda_1, \ldots, \lambda_P]'$ is the parameter vector, the $N \times P$ matrix of known regressors $\mathbf{G}$ describes the deterministic part of the system, and the random process $\{u_n\}$ represents the intrinsic randomness of the phenomena, with $\mathbf{u} := [u_1, \ldots, u_N]'$. This framework includes many models of current importance for system analysis, for example the Fourier, Wavelet and other expansions used in Compressive Sensing, Communication Systems, Genomics and Signal Processing. In particular, when $\mathbf{G}$

[1] Note, however, that many real-world systems are intrinsically non-linear and continuous in time. Most of the analysis that will be developed is directly applicable to non-linear models, while we address the continuous-time scenario in the final chapter of this dissertation.

contains past realizations of $\mathbf{x}$, of the form[2]

$$\mathbf{G} = \begin{bmatrix} x_0 & x_{-1} & \dots & x_{-(P-2)} & x_{-(P-1)} \\ x_1 & x_0 & \dots & x_{-(P-3)} & x_{-(P-2)} \\ \vdots & & \ddots & & \vdots \\ x_{N-2} & x_{N-3} & \dots & x_{N-P} & x_{N-P-1} \\ x_{N-1} & x_{N-2} & \dots & x_{N-P+1} & x_{N-P} \end{bmatrix}, \tag{1.2}$$

then (1.1) becomes the standard autoregressive model of order P, AR(P).

We may further assume that the states x_n are observed in additive noise,

$$y_n = bx_n + v_n, \qquad n = 1, \dots, N, \tag{1.3}$$

where b is an observation parameter and $\{v_n\}$ is the observation noise process. Then the system of equations (1.1) and (1.3) can be put in the form of a discrete-time, linear, state-space model (SSM),[3] that has a cardinal role in time series modelling.

Depending on the application at hand, there are many possible inference objectives; for example, state inference or prediction for x_n in the SSM, parameter estimation for $\boldsymbol{\lambda}$ and b, and model choice to determine the structure and dimensionality (P) of the model. A common assumption that facilitates any inference task is to have u_n and v_n normally distributed, an hypothesis that has, firstly, a well grounded theoretical justification. In fact, the driving noise u_n can often be thought of as the sum of $m \gg 1$ small contributions, of the form

$$u_n := a_m^{-1}(T_{1n} + \dots + T_{mn} - d_m) \tag{1.4}$$

where $T_{ij} \sim F$, $i = 1, \dots, m$, $j = 1, \dots, n$, are independent and identically distributed (i.i.d.) random variables (RVs), $a_m > 0$ is a scaling constant and $d_m \in \mathbb{R}$ is a centering constant. According to the most widely known version of the classical central limit theorem (CLT), if

$$\mathbb{E}[T_{ij}^2] < \infty, \qquad a_m = (m \operatorname{Var}[T_{ij}])^{1/2}, \qquad d_m = m\mathbb{E}[T_{ij}], \tag{1.5}$$

then u_n converges weakly, or in distribution, to the standard Gaussian distribution, as $m \to \infty$, see e.g. [Feller, 1966, p. 515]. Similarly, the measurement noise

[2] The first row of the matrix $\mathbf{G}$ requires initialization of the values $x_0, \dots, x_{-(P-1)}$, while the consecutive rows are progressively computed, as new realizations of the process become available.

[3] After specifying the state transition density for x_n.

process $\{v_n\}$ typically consists of sums of small perturbations, a situation again well-described by the Gaussian distribution on v_n. In this setting, simple and standard methods are available for likelihood-based or Bayesian inference, for the tasks mentioned above, using closed-form results combined with, for example, Variational Bayes (including Expectation Maximization (EM)) or Monte Carlo sampling.

However, many real-world cases exhibit extreme values much more frequently than the Gaussian model for $\{u_n\}$ and $\{v_n\}$ in (1.1) would allow. Examples of such abrupt changes include variations presented by stock prices or insurance gains/losses in financial applications, as studied extensively since the seminal works of Mandelbrot [1963] and Fama [1965], and that are constantly at the center of financial modelling, see Rachev and Mittnik [2000] for a recent overview. Further applications can be found in various fields of engineering, such as communications (see Azzaoui and Clavier [2010] for statistical modelling of channels, de Freitas et al. [2017] for capacity bounds, Liebeherr et al. [2012] for delay bounds in networks with α-stable noise, and Shevlyakov and Kim [2006]; Warren and Thomas [1991] for signal detection), signal processing Nikias and Shao [1995], image analysis Achim et al. [2001, 2006] and audio processing Lombardi and Godsill [2006]. Sudden changes in the observed phenomena are studied also in the climatological sciences, see Katz and Brown [1992]; Katz et al. [2002], and in the medical sciences, see, e.g., Chen et al. [2010] on brain connectivity representations. Moreover, in the field of sparse modelling and Compressive Sensing, a noise distribution is required that leads to sparse solutions in transformed domains, a case much better dealt with using heavy-tailed models than the Gaussian case; see, e.g., Amini and Unser [2014]; Carrillo et al. [2016]; Tzagkarakis [2009]; Unser and Tafti [2014].

One of the simplest ways to account for such large (extreme) observations in our motivational linear state-space model is to attribute a *heavy-tailed* distribution to u_n in (1.1), while a Gaussian distribution for v_n in (1.3) could be kept to signify that the measurements y_n do not depart too much from the "true" state of the system x_n.[4] A distribution is said to be heavy-tailed,[5] see e.g. Simon [1955], Mitzenmacher [2004], Sornette [2006], Clauset et al. [2009], Stumpf and Porter [2012], if its probability

[4] However, modelling u_n as heavy-tailed is just as important as modelling v_n as such, in certain applications.

[5] The class of heavy-tailed distributions is actually larger than the one that we consider here, and that forms the sub-class of fat-tailed distributions. In fact, heavy-tailed distributions include all those probability laws that have tail decay that is not exponentially bounded. According to the rate of this decay, one can define the sub-classes of subexponential, long-tailed, or fat-tailed distributions, among others.

density function (PDF), $p(x)$, has power-law tail decay

$$\lim_{|x|\to\infty} \frac{p(x)}{|x|^{-1-\alpha}} = c, \tag{1.6}$$

or, equivalently, $p(x) \sim c|x|^{-1-\alpha}$, for $|x| \to \infty$,[6] where $\alpha > 0$ is a parameter of the distribution, and $c > 0$ is a constant, that can depend on α, as well as on other distribution parameters, and the sign of x. When (1.6) holds only for $x \to +\infty$ or $x \to -\infty$, the distribution is heavy-tailed on one side only. Many distributions in common use present tail decay as in (1.6), including the Student's t, the Pareto, the log-logistic, the Cauchy and the Lévy distributions. The last two examples (Cauchy and Lévy distributions) belong to the class of statistical distributions that is central to this dissertation, the class of (non-Gaussian) α-stable distributions.

The main motivation for our focus on this class, as well as the main driving force for the large attention that the α-stable laws have received in applications (see the extensive bibliography listed in Nolan [2018a]), both stem from the key role of the α-stable distribution in the generalized CLT, and from the modelling flexibility offered by this class of laws.

1.2 Generalized Central Limit Theorem and Stable Distributions

Even if heavy-tailed, the random process $\{u_n\}$ considered in many of the situations listed in the above section can be thought of as emerging from the combination of many independent perturbations of the form (1.4). Whenever the centred and normed sum of i.i.d. RVs converges in distribution, it converges to a member of the class of α-stable distributions, with $\alpha \in (0, 2]$, according to the generalized CLT [Gnedenko and Kolmogorov, 1968, p. 162],[Feller, 1966, p. 576], [Embrechts et al., 1997, p. 79]. The Gaussian is a special member of this class (recovered for $\alpha = 2$), the only one with finite variance. Hence, using non-Gaussian α-stable distributions for the elements of the vector $\mathbf{u}$ in (1.1) offers a way of modelling time series with large (extreme) values.

When (1.4) converges in distribution, then the summands T_{ij}, are said to *belong to the domain of attraction*, DA(α), of the α-stable distribution. When the scaling coefficients can be chosen to be of the form $a_m = km^{1/\alpha}$, for some positive constant k,

[6] Throughout, we use the symbol '$\sim$' to denote the fact that a RV $X \sim \mathcal{D}$ has distribution $\mathcal{D}$, but also to denote the following asymptotic relationship: for two (real) functions f_1 and f_2, we say $f_1(x) \sim f_2(x)$ when $\lim_{x\to\infty} f_1(x)/f_2(x) = 1$.

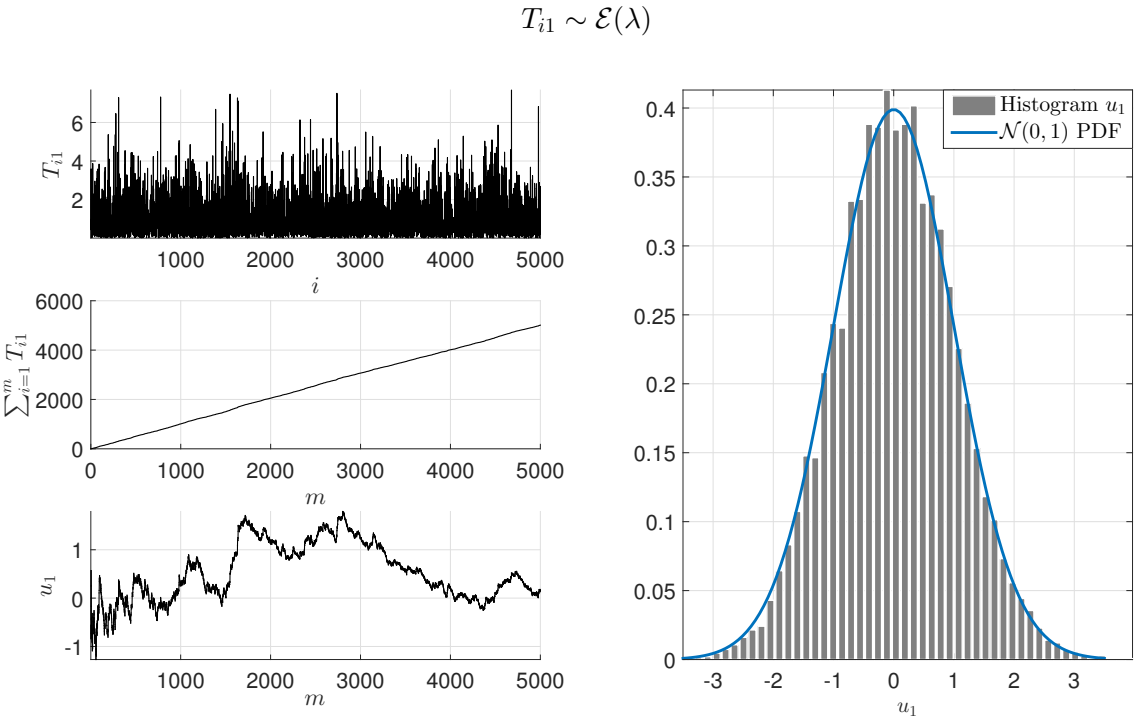


Figure 1.1 Top-left: one sample path from the sequence $\{T_{i1}\}_{i=1}^{5\times10^3}$, with $T_{i1} \sim \mathcal{E}(\lambda = 1)$; centre-left: corresponding sample path of the cumulative sum $\sum_{i=1}^{m} T_{i1}$; bottom-left: corresponding sample path of the centred sum u_1 defined as in (1.4), with $a_m = \sqrt{m}$, $d_m = m$; right: normalized histogram of 10^4 variables u_1, compared to the standard Gaussian PDF.

then the variables T_{ij} are said to *belong to the domain of normal attraction*, DNA(α), of the α-stable distribution. The conditions (1.5) in the classical CLT are related to DNA(2), the domain of normal attraction of the Gaussian distribution. However, the Gaussian distribution also attracts variables with infinite variance, see e.g. the Lindeberg's CLT Lindeberg [1922], [Gnedenko and Kolmogorov, 1968, p.102, 103],[Feller, 1966, p.262, 518]. We refer, more generally, [Gnedenko and Kolmogorov, 1968, p.172, 181], [Feller, 1966, p.312-313, 577-581], [Embrechts et al., 1997, p. 79] for the necessary and sufficient conditions for convergence to the Gaussian or non-Gaussian α-stable distributions.

Figure 1.1 illustrates, for example, how the Exponential distribution with mean $1/\lambda$, $\mathcal{E}(\lambda)$, belongs to DNA(2), being attracted by the Gaussian distribution, while Figure 1.2 shows how the Lévy(μ, σ) distribution with location μ and scale σ belongs to DNA(0.5). In particular, each α-stable distribution belongs to DNA(α), corresponding, in the example in Figure 1.2, to the fact that the Lévy distribution is a member of the α-stable class, with $\alpha = 0.5$.[7]

[7]The location and scale parameters of the attracting Lévy distribution in the example in the Figure 1.2 have been computed according to [Roy, 2012, p.35], resulting in $\mu = 0$ and $\sigma = 1.57$.

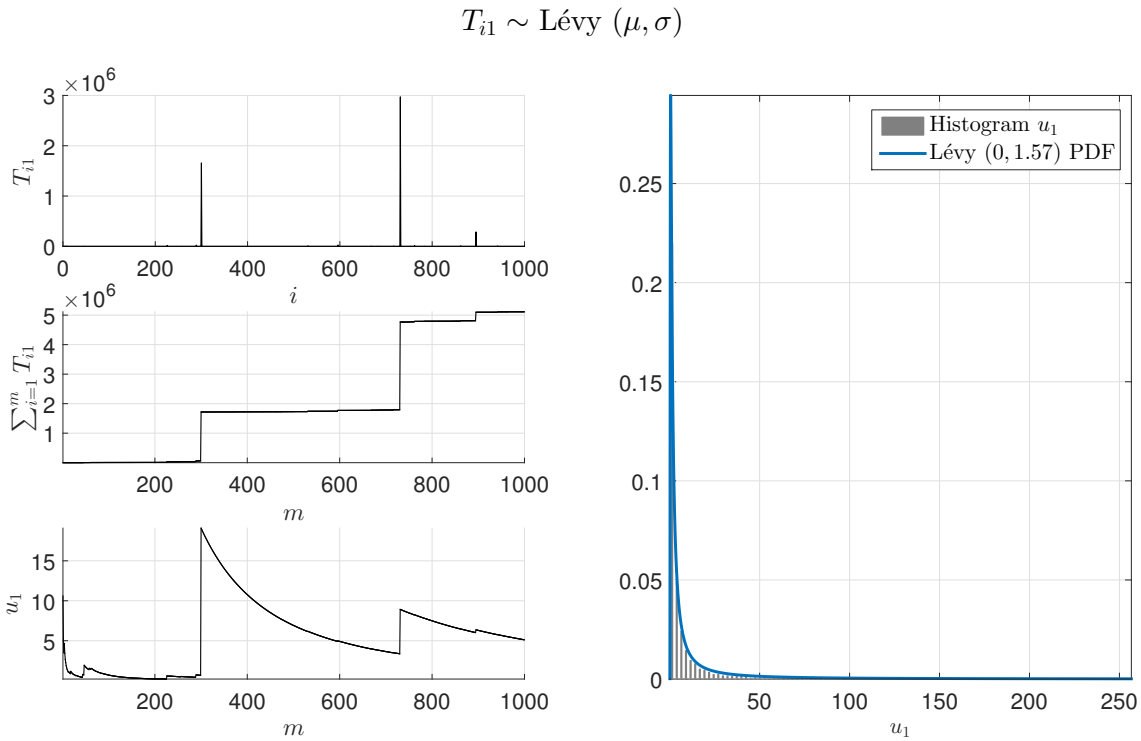


Figure 1.2 Top-left: one sample path from the process $\{T_{i1}\}_{i=1}^{10^3}$, with $T_{i1} \sim$Lévy$(\mu = 10, \sigma = 1)$; centre-left: corresponding sample path of the cumulative sum $\sum_{i=1}^{m} T_{i1}$; bottom-left: corresponding sample path of the centred sum u_1 defined as in (1.4), with $a_m = m^2$, $d_m = 0$; right: normalized histogram of 10^4 variables u_1, compared to the Lévy$(\mu = 0, \sigma = 1.57)$ PDF.

The generalized CLT is a constructive definition of the α-stable distribution, as the only infinitely divisible distribution[8], that is the weak limit of (1.4). Alternatively, the α-stable distribution is defined through its characteristic function (CF) $\phi_X(s) := \mathbb{E}\left[\exp\left(isX\right)\right]$, for $s \in \mathbb{R}$, detailed in Sections 2.1 and 2.2 for some of the commonly used parametrizations, together with the meaning of the distribution parameters.

Unlike the CF, the PDF of α-stable distributions cannot be expressed in closed form, except in a few special cases. We refer to [Feller, 1966, p. 581] and [Uchaikin and Zolotarev, 1999, Chapter 4] for series expansions of the stable PDF, Hoffmann and Jørgensen [1994] and Zolotarev [1995] for representation through special functions, and to Nolan [1997] and Ament and O'Neil [2018] for examples of numerical schemes for evaluation of the PDF. This lack of simple expressions of the PDF is a significant complication when developing effective tools for inference about α-stable distributions. Nevertheless, a wide variety of statistical methods have been proposed in the literature,

[8] An infinitely divisible distribution is one that can be represented as the sum of n RVs with common distribution F_n, for each $n \in \mathbb{N}$, see e.g. [Feller, 1966, p.176] or [Gnedenko and Kolmogorov, 1968, p. 115]. The Gamma and the Poisson distributions are examples of infinitely divisible distributions.

Table 1.1 Summary of common inference categories and of the choice of target in this thesis, in terms of methodological and analytical development.

Category of Inference	The Target of this Thesis
The framework: frequentist versus Bayesian.	We address Bayesian inference methods, given that they naturally allow for uncertainty quantification, which is important in the heavy-tails scenario implicit when using stable distributions. Furthermore, the Bayesian framework is well suited for LV models.
The nature: exact versus approximate.	Given the lack of closed form PDF (or likelihood), many inference methods for stable distributions are approximate. However some of the LV models for the stable distribution allow for exact inference, and we contribute towards these. When the chosen LV model requires approximations, we aim at quantifying the amount of error introduced.
The underlying distribution or stochastic process (scalar or multivariate), and its parameter regime.	We consider mostly scalar, both symmetric and skewed stable RVs, However, we provide some results on the simulation of multivariate stable stochastic processes that are the solution to stochastic differential equations (SDEs); this is a first step towards future development of inference methods.
The aim: learning the distribution parameters, or inferring the state of the system.	We consider batch, distribution and model, parameter learning problems, but the LV models that we examine also allow for sequential state inference.

motivated by the central role of this class; we refer again to Nolan [2018a] for an extensive bibliography.

1.3 Inference with Stable Distributions

For systems governed by α-stable noise, many inference examples can be found in the references to specific application domains (finance, engineering, climatological and medical sciences) provided in Section 1.1. For the sake of clarity, we focus mostly on ways for estimating the parameters $\boldsymbol{\lambda}$ in our motivating example (1.1) above, together with the distribution parameters. These coincide with the vector $\boldsymbol{\theta} = (\alpha, \sigma, \beta, \mu)$ for scalar stable RVs, or with a scalar α, a vector $\boldsymbol{\mu}$ and a function ξ for stable random vectors, as detailed in Chapter 2.

Numerous techniques have been developed for estimating the parameter vector $\boldsymbol{\theta}$ in the scalar case. Common frequentist approaches include, among others, those based on the quantiles of the distribution McCulloch [1986], its logarithmic moments Kuruoğlu [2001], the empirical CF Paulson et al. [1975] and Koutrouvelis [1980], approximate maximum likelihood estimators Mittnik et al. [1999]; Nolan [2001], or block-maxima scaling Stoev et al. [2011]. However, since the α-stable PDF is not available in closed form, all the above approaches are approximate. This issue similarly affects corresponding Bayesian methods aiming at computing the posterior distribution of the parameters, for which the likelihood function needs to be evaluated. We refer, for example, to the posterior sampler of Lombardi [2007], based on numerical inverse-Fourier transforms of the CF. A similar scenario repeats for inference of the parameters of multivariate stable distributions. Approximate methods are available, with the further drawback that, in order to avoid inference of the infinite-dimensional function ξ, a simplified structure for the multivariate distribution is usually assumed (e.g. discrete ξ, or sub-Gaussianity of the stable vector, as explained in Chapter 2). The frequentist inference schemes of Nolan et al. [2001] and Nolan [2013] are based on the empirical CF and one-dimensional projections of the data, while approximate Bayesian samplers based on summary statistics are provided by Peters et al. [2011] and Peters et al. [2012].

On the other hand, the α-stable distribution admits representations[9] involving latent variables, that allow for (asymptotically) exact or nearly exact Bayesian inference, and circumvent the lack of closed form expressions for the likelihood. Our work focuses on such representations, summarized in Chapters 2 and 3, where the choice of the Bayesian framework for inference is also and primarily motivated by the ability of providing complete (posterior) distributions of the parameters of interest, an important feature in our heavy tails scenario, as explained in Section 3.1.

A first latent variables model derives from the method for generating stable RVs proposed by Chambers et al. [1976], and it has enabled the conditional samplers of Buckle [1995] and Qiou and Ravishanker [1998a]. Such inference methods are nominally exact; however, numerical steps need to be introduced to invert highly non-linear functions, and that destroy the exactness of the schemes. Based on the same LV model, and the pseudo-marginal MCMC inference scheme of Andrieu and Roberts [2009], our first contribution is the development of an exact and efficient pseudo-marginal sampler, as detailed in Chapter 4.

Alternatively, the product property [Feller, 1966, p. 176] and the scale mixture of normals (SMiN) representation of symmetric stable distributions have been used in

[9] Given two RVs X and Y, in the following we say that X has the *representation* Y if $X \overset{\mathcal{D}}{=} Y$, where $\overset{\mathcal{D}}{=}$ denotes equality in distribution.

Godsill [1999, 2000]; Godsill and Kuruoğlu [1999]; Tsionas [1999]. Such inference methods are exact, but are valid only for symmetric stable distributions, and can suffer from poor convergence for extremely large or small data values. Numerical approximations are proposed in the literature to improve convergence, but they introduce inaccuracies that are difficult to assess.

A generalization of the SMiN scheme to the skewed stable distribution is given by the mean and scale mixture of normals (MaSMiN) representation that derives from the Poisson series representation (PSR). As summarized in Chapter 3, the Bayesian samplers of Lemke [2014]; Lemke and Godsill [2014]; Lemke et al. [2015] are based on the MaSMiN model. However, also these schemes are approximate: in fact, although the PSR provides an exact representation, this is in the form of an infinite series that needs to be approximated. Our main contribution in Chapter 5 is to formally justify the approximation used in the above literature through a CLT for the PSR residual (the remainder of the series), that is thus approximated with a Gaussian RV. Given that our main drive is aiming for exact (or nearly exact) inference techniques, we believe that, for effective inference, it is necessary to quantify the error incurred by such approximations. One of the central contributions of this thesis is error quantification for the PSR of symmetric stable distributions, as presented in Chapters 6 and 7. This is a first step towards error quantification for the PSR for a generally skewed RV, to be developed in future work.

Multivariate stable distributions also admit a PSR. Remarkably, such representation can introduce a re-parametrization of the multivariate parameters from $\alpha, \boldsymbol{\mu}, \xi$ to $\alpha, \boldsymbol{\mu}_W, \boldsymbol{\Sigma}_W$, where $\boldsymbol{\mu}_W$ is a vector, and $\boldsymbol{\Sigma}_W$ is a matrix. A similar re-parametrization is maintained also after the approximation of the PSR, an approximation necessary in analogy with the scalar case. This translates into the fact that the complexity of the parameter inference problem for a general multidimensional stable distribution can be greatly reduced, without the need of assuming discrete ξ or sub-Gaussian structure, as in the inference schemes mentioned above.

In this thesis we do not implement such inference schemes, leaving this to future work. We rather focus on formulating the approximation of the multivariate PSR. We do so in the context of (approximate) simulation from linear multivariate stochastic differential equations (SDEs) driven by stable noise processes. These can be considered a continuous-time extension of (1.1), and are detailed in Section 2.3.2. Simulating the multivariate stable distribution that corresponds to the solution of such SDEs is the key step for implementing Bayesian state inference methods such as sequential Monte-Carlo (SMC), Cappé et al. [2007]; Doucet et al. [2000]; Doucet and Johansen [2011]. When part of the state is (conditionally) linear and Gaussian, combining Kalman-filter steps Kalman [1960] with SMC filters results in

more efficient filters in terms of Monte Carlo variance, Doucet et al. [2000]; Schön et al. [2005]. Both the SMiN and the MaSMiN representations of the stable noise in linear dynamics imply that the state is conditionally Gaussian, and allow to implement such ideas. The SMiN representation is used by Lombardi and Godsill [2004] and Lombardi and Godsill [2006] for a d-dimensional state, while Lemke and Godsill [2011], Lemke [2014], Lemke and Godsill [2015] rely on the MaSMiN for scalar state. The extension of the latter is our drive for the adaptation of the PSR and its approximation to stable linear models with multidimensional state in Chapter 8. The development of state inference techniques, as well as the quantification of the amount of approximation introduced is left to future developments.

Table 1.1 gives a summary of some common categorizations of inference methods, and it specifies those that this thesis is concerned with and relevant for.

1.4 Main Contributions and Thesis Organisation

The structure of this dissertation is as follows: Chapters 2 and 3 contain background material on α-stable distributions and processes, and on Baysian parameter inference, with emphasis on the above mentioned LV models. Our contributions are presented in Chapters 4 - 8. In particular:

- Chapter 4 provides an exact and efficient pseudo-marginal sampler for Bayesian inference of the scalar distribution parameters in the generally skewed case;
- Chapter 5 contains the proof of the asymptotic normality of the PSR residual, and the expressions of some relevant CFs (PSR residual and truncated PSR); *part of the work in this chapter has been developed in collaboration with Dr. Tohid Ardeshiri and Prof. Ioannis Kontoyiannis*;
- Chapter 6 presents nonasymptotic numerical and analytic bounds on the Komogorov distance between the PSR residual and the Gaussian distribution; *the work in this chapter has been developed in collaboration with Prof. Ioannis Kontoyiannis*;
- Chapter 7 presents nonasymptotic numerical and analytic bounds on the Kolmogorov distance between the truncated PSR with added Gaussian approximation of the residual and the stable distribution;
- Chapter 8 provides an approximation of the residual of the PSR for multivariate linear SDEs driven by stable noise, and it shows approximate simulation from the model.

Finally, Chapter 9 contains a a summary and a discussion of the work done, by presenting its limitations, implications, and insights on future extensions.

1.5 Publications

Parts of this thesis have been/are going to be published in the following papers.

Submitted

- **Riabiz**, M., Ardeshiri, T., Kontoyiannis, I., and Godsill, S., *Nonasymptotic Gaussian Approximation for Linear Systems with Stable Noise*. arXiv:1802.10065v3. Preprint submitted to IEEE Transactions on Information Theory.

Peer Reviewed Journal Publications

- Lemke, T., **Riabiz**, M., and Godsill, S., *Fully Bayesian Inference for α-Stable Distributions Using a Poisson Series Representation*. Digital Signal Processing, Volume 47, Issue C, December 2015, Pages 96-115.

Peer Reviewed Conference Publications

- **Riabiz**, M., Ardeshiri, T., Kontoyiannis, I., and Godsill, S., *Sharp Gaussian approximation bounds for linear systems with α-stable noise*. Proceedings of the IEEE International Symposium on Information Theory, 2018, Vail, Colorado.
- **Riabiz**, M., Ardeshiri, T., Kontoyiannis, I., and Godsill, S., *Simulated convergence rates with application to an intractable α-stable inference problem*. Proceedings of the 7th IEEE workshop on Computational Advances in Multi-Sensor Adaptive Processing, 2017, Curaccao, Dutch Antilles.
- **Riabiz**, M., Ardeshiri, T., and Godsill, S., *Convergence results for tractable inference in α-stable stochastic processes*. Proceedings of the 22nd International Conference on Digital Signal Processing, 2017 (DSP '17), London, UK. Pages 1-5.
- **Riabiz**, M., and Godsill, S., *Approximate simulation of linear continuous time models driven by asymmetric stable Lévy processes*. Proceedings of the 42nd International Conference on Acoustics, Speech, and Signal Processing, 2017 (ICASSP '17), New Orleans, Louisiana. Pages 4676 - 4680.

- **Riabiz**, M., Ardeshiri, T., and Godsill, S., *A central limit theorem with application to inference in α-stable regression models*. Proceedings of the Time Series Workshop at NIPS 2016, Barcelona, Spain, and PMLR 55:70-82, 2017.

- **Riabiz**, M., Lindsten F., and Godsill, S., *Pseudo-Marginal MCMC for Parameter Estimation in α-Stable Distributions*. In Proceedings of the 17th IFAC Symposium on System Identification (SYSID), 2015, Beijing, China. Volume 48, Issue 28, 2015, Pages 472-477

Chapter 2

Background I: Stable Distributions

This Chapter first provides insight on scalar and multivariate α-stable distributions, as well as on stable Lévy processes, and stochastic integrals that are at the solution of linear SDEs driven by stable noise. We then give the details of LV models available for stable distributions, specifying the parameter regimes for which they hold. Such LV models are at the basis of the inference techniques summarized in Chapter 3, and justify the studies contained this thesis.

2.1 Univariate Stable Distributions

We adopt the standard notation of Samoradnitsky and Taqqu [1994] and we recall some equivalent definitions and properties of α-stable distributions on $\mathbb{R}$. The notion of 'stability' implicit in the name of α-stable distributions is related to the first characterization of such a class, given by Lévy [1924], Lévy [1925], Khinchine and Lévy [1936]: stable distributions are preserved under convolution. In the following, $\stackrel{\mathcal{D}}{=}$ and $\stackrel{\mathcal{D}}{\longrightarrow}$ denote equality and convergence in distribution, respectively.

Definition 2.1.1. (Stability, convolution) A RV X is said to have stable distribution if, $\forall a_1 > 0, a_2 > 0$, there exist $a_3 > 0$, and $d \in \mathbb{R}$ such that

$$a_1 X_1 + a_2 X_2 \stackrel{\mathcal{D}}{=} a_3 X + d, \tag{2.1}$$

where X_1 and X_2 are independent copies[1] of X.

When in (2.1) we can choose $d = 0$, the variable X is said to be *strictly stable*, a case that includes all *symmetric* stable distributions (meaning that X and $-X$ have

[1]The term *independent copies* is borrowed from Samoradnitsky and Taqqu [1994] and it is equivalent to saying that X_1 and X_2 are i.i.d., and have the same distribution as X.

the same law). Any strictly stable distribution admits a real number $\alpha \in (0, 2]$ such that

$$a_1^\alpha + a_2^\alpha = a_3^\alpha,$$

see [Feller, 1966, p. 171]. As anticipated in Chapter 1, α is one of the parameters of the stable distribution, the one that governs the asymptotic power-law tail decay of the form (1.6).

The second definition of stability that we give generalizes (2.1) to an arbitrary number of summands, and it is related to the fact that the α-stable distribution belongs to its own domain of normal attraction in the generalized CLT.

Definition 2.1.2. (Stability, α-stable in DNA(α)) A RV X is said to have stable distribution if, for any $n \geqslant 2$, there are $a_n > 0$ and $d_n \in \mathbb{R}$ such that

$$a_n^{-1}(X_1 + X_2 + \ldots + X_n - d_n) \stackrel{\mathcal{D}}{=} X, \tag{2.2}$$

where $X_1, X_2, \ldots, X_n$ are independent copies of X. Specifically, if (2.2) holds, then $a_n = n^{1/\alpha}$, for some $\alpha \in (0, 2]$.

However, as anticipated in Chapter 1, the stable distribution *attracts* a large number of other distributions, according to the generalized CLT, as in the following definition.

Definition 2.1.3. (Stability, generalized CLT) A RV X is said to have a stable distribution if it has a domain of attraction, that is if there is a sequence of i.i.d. random variables T_1, T_2, $\ldots T_n$ and sequences $a_n > 0$, $d_n \in \mathbb{R}$, such that

$$a_n^{-1}(T_1 + T_2 + \ldots + T_n - d_n) \xrightarrow[n\to\infty]{\mathcal{D}} X.$$

In the above definition, if $\mathrm{Var}[T_i] < \infty$, then $a_n = (n \,\mathrm{Var}[T_i])^{1/2}$ and X is the standard normal distribution, with mean 0 and variance 1. If $\mathrm{Var}[T_i] = \infty$, then $a_n = n^{1/\alpha} L(n)$, for $\alpha \in (0, 2]$, and L a slowly varying function at infinity, that is, $\lim_{x\to\infty} L(ux)/L(x) = 1$ for all $u > 0$, and X is either Gaussian or non-Gaussian α-stable, [Feller, 1966, p. 576].

Finally, we can define the stable distribution by providing its CF. Specifying the CF corresponds to choosing one of the possible parametrizations of the stable law. The two main parametrizations in use are the Zolotarev [1986]'s A, or $\mathcal{S}^1$, formulation and the M, or $\mathcal{S}^0$, formulation. We refer to [Samoradnitsky and Taqqu, 1994, p. 5-9], Nolan [1997] and Nolan [1998] for a detailed analysis and comparison of the properties of the two. We focus on the $\mathcal{S}^1$ parametrization, that provides a

natural interpretation of the parameters of the distribution, see also [Gnedenko and Kolmogorov, 1968, p. 164].

Definition 2.1.4. (Stability, CF $\mathcal{S}^1$) A random variable X is said to have stable distribution iff its CF, $\phi_X(s)$, is such that

$$\log \phi_X(s) = \begin{cases} -\sigma^\alpha |s|^\alpha \left\{1 - i\beta \operatorname{sgn}(s) \tan \frac{\pi\alpha}{2}\right\} + i\mu s, & \alpha \neq 1, \\ -\sigma|s| \left\{1 + i\beta \operatorname{sgn}(s) \frac{2}{\pi} \log |s|\right\} + i\mu s, & \alpha = 1. \end{cases} \tag{2.3}$$

where $\alpha \in (0, 2]$, $\beta \in [-1, 1]$, $\sigma > 0$, $\mu \in \mathbb{R}$.

Then we write $X \sim \mathcal{S}^1_\alpha(\sigma, \beta, \mu)$, or simply $X \sim \mathcal{S}_\alpha(\sigma, \beta, \mu)$ in the following, to indicate that the X has stable distribution with parameters α, σ, β and μ. The first parameter, $\alpha \in (0, 2]$, is the tail parameter, given that, when $\mu = 0$ and $\alpha \in (0, 2)$, the PDF $p(x)$ of X has tails that decay (1.6), with more extreme values (and hence heavier tails) appearing more frequently for smaller values of α. The second parameter, $\beta \in [-1, 1]$, is a measure of skewness: $\beta = 0$ corresponds to *symmetric* stable laws, while $\beta = \pm 1$ corresponds to the fully left or right skewed cases.[2] Finally, $\mu \in (-\infty, \infty)$ and $\sigma > 0$ are the location and scale parameters, respectively, with *standard stable* distributions corresponding to $\sigma = 1$ and $\mu = 0$, and *strictly stable* distributions to $\mu = 0$.[3] Observe that μ is not the mean of the distribution, since it is finite only if $\alpha \leq 1$. Similarly, σ is not the standard deviation, the variance of a stable distribution being finite only if $\alpha = 2$.[4] The infinite variance property has generated a debate about the applicability of the stable law, given that it is not possible to assess the infinite variance of observations; see, for example, Grabchak and Samorodnitsky [2010] for a discussion related to modelling of financial datasets. Nonetheless, we pursue our analysis of stable models, motivated by the fact that phenomena with very large variance are observed in practice, as well as by the generalized CLT argument.

Three specific choices of the parameters lead to the only cases in which the distribution admits a closed form PDF: $\alpha = 2$ corresponds to the Gaussian case, with variance $2\sigma^2$ and mean μ; $\alpha = 1$, $\beta = 0$ corresponds to the Cauchy distribution, with scale σ and location μ; $\alpha = 0.5$, $\beta = \pm 1$ is the Lévy distribution with scale σ and location μ.

[2] When $\beta = \pm 1$, the distribution is single-tailed only when $\alpha < 1$, see [Samoradnitsky and Taqqu, 1994, p. 13], while it has support on the real line if $\alpha \in [1, 2]$.

[3] Observe that $\beta = 0$ implies $\mu = 0$, meaning that all symmetric stable distributions are strictly stable.

[4] In general, an α-stable RV X with $0 < \alpha < 2$ has finite absolute moments $\mathbb{E}[|X|^p]$, with $0 < p < \alpha$, see [Samoradnitsky and Taqqu, 1994, p.18]. When α=2, moments of all the orders are finite.

Remark 1. Version of the $\mathcal{S}^1$ parametrization. A version of the $\mathcal{S}^1$ formulation (2.3), whose use can be justified by analytic considerations (see Zolotarev [1986], Weron [1996a], [Samoradnitsky and Taqqu, 1994, p. 8]), and that we use in Chapter 4, has CF $\tilde{\phi}_X(s)$ as follows

$$\log \tilde{\phi}_X(s) = \begin{cases} -\sigma_2^\alpha |s|^\alpha \left\{ -i\beta_2 \operatorname{sgn}(s) \frac{\pi}{2} K(\alpha) \right\} + i\mu s, & \alpha \neq 1, \\ -\sigma_2 |s| \left\{ \frac{\pi}{2} + i\beta_2 \operatorname{sgn}(s) \log |s| \right\} + i\mu s, & \alpha = 1, \end{cases} \tag{2.4}$$

where

$$K(\alpha) := \alpha - 1 + \operatorname{sgn}(1-\alpha) = \begin{cases} \alpha, & 0 < \alpha < 1, \\ \alpha - 2, & 1 < \alpha < 2. \end{cases} \tag{2.5}$$

In this case, we write $X \sim \mathcal{S}^1_\alpha(\sigma_2, \beta_2, \mu)$, or simply $X \sim \mathcal{S}_\alpha(\sigma_2, \beta_2, \mu)$. The transformation between the parameters (σ_2, β_2) and (σ, β) is given by

$$\beta = \begin{cases} \dfrac{\tan\left(\frac{\beta_2 \pi K(\alpha)}{2}\right)}{\tan\left(\frac{\pi\alpha}{2}\right)}, & \alpha \neq 1, \\ \beta_2, & \alpha = 1; \end{cases} \qquad \sigma = \begin{cases} \dfrac{\sigma_2}{\left(1 + \beta^2 \tan^2\left(\frac{\pi\alpha}{2}\right)\right)^{\frac{1}{2\alpha}}}, & \alpha \neq 1, \\ \frac{\pi}{2}\sigma_2, & \alpha = 1. \end{cases} \tag{2.6}$$

Remark 2. Relation between $\mathcal{S}^0$ and $\mathcal{S}^1$ parametrizations. The CFs (2.3) and (2.4) are discontinuous at all points $\alpha = 1$, $\beta \neq 0$. This led to the introduction of the $\mathcal{S}^0_\alpha(\sigma, \beta, \mu_0)$ parametrization, in which the new location parameter $\mu_0 = f(\mu)$ makes the new CF jointly continuous in α and β, see e.g. [Samoradnitsky and Taqqu, 1994, p. 8] or Nolan [1998]. However μ_0 loses the natural interpretation of location parameter and we prefer referring to the $\mathcal{S}^1$ parametrization in the rest of this thesis, assuming that $\alpha \neq 1$ for the sake of simplicity.

2.1.1 Simulation of Stable Random Variables

Chambers et al. [1976] presented a method (which we refer to as the CMS method) for generating stable RVs, that consists of (highly) non-linear transformations applied to an exponential and a uniform RV. We give the details for the two versions of the $\mathcal{S}^1$ parametrization introduced above. In our notation $\mathcal{U}(a, b)$ indicates the continuous uniform distribution on the interval (a, b), while $\mathcal{E}(\lambda)$ denotes the exponential distribution with mean $1/\lambda$.

CMS Method for Simulating $\mathcal{S}_{\boldsymbol{\alpha}}(\boldsymbol{\sigma}, \boldsymbol{\beta}, \boldsymbol{\mu})$

Let us define the function $g_{\alpha,\beta}(y)$, for $y \in (-\pi/2, \pi/2)$,

$$g_{\alpha,\beta}(y) := s_{\alpha,\beta} \frac{\sin(\gamma_{\alpha,\beta} + \alpha y)}{(\cos y)^{1/\alpha}} \left(\cos(\gamma_{\alpha,\beta} + (\alpha - 1)y)\right)^{\frac{1-\alpha}{\alpha}}, \tag{2.7}$$

where the constants $\gamma_{\alpha,\beta}$ and $s_{\alpha,\beta}$ are defined as follows

$$\gamma_{\alpha,\beta} := \arctan\left(\beta \tan\left(\frac{\pi\alpha}{2}\right)\right), \qquad s_{\alpha,\beta} := \frac{1}{(\cos(\gamma_{\alpha,\beta}))^{1/\alpha}} = \left(1 + \beta^2 \tan^2\left(\frac{\pi\alpha}{2}\right)\right)^{\frac{1}{2\alpha}}.$$

Then the following theorem provides a way of generating $\mathcal{S}_\alpha(\sigma, \beta, \mu)$ RVs.

Theorem 2.1.1. Let $\alpha \in (0, 2]$ and $\beta \in [-1, 1]$ and let $g_{\alpha,\beta}(y)$ be defined as in (2.7). Let $Y \sim \mathcal{U}(-\pi/2, \pi/2)$ and $W \sim \mathcal{E}(1)$, independent of Y. Let

$$Z_1 := \begin{cases} g_{\alpha,\beta}(Y)\, W^{\frac{\alpha-1}{\alpha}}, & \alpha \neq 1, \\ \frac{2}{\pi}\left[\left(\frac{\pi}{2} + \beta Y\right)\tan Y - \beta \log\left(\frac{\frac{\pi}{2} W \cos Y}{\frac{\pi}{2} + \beta Y}\right)\right], & \alpha = 1. \end{cases}$$

Then $Z_1 \sim \mathcal{S}_\alpha(1, \beta, 0)$, with CF given by (2.3).

CMS Method for Simulating $\mathcal{S}_{\boldsymbol{\alpha}}(\boldsymbol{\sigma_2}, \boldsymbol{\beta_2}, \boldsymbol{\mu})$

A first way of generating stable RVs in the $\mathcal{S}_\alpha(\sigma_2, \beta_2, \mu)$ parametrization is by combining the above CMS method for $\mathcal{S}_\alpha(\sigma, \beta, \mu)$ with the transformations (2.6) on the parameters. Alternatively, let us define the function $t_{\alpha,\beta_2}(y)$, for $y \in (-1/2, 1/2)$,

$$t_{\alpha,\beta_2}(y) := \frac{\sin(\pi\alpha y + \eta_{\alpha,\beta_2})}{(\cos(\pi y))^{1/\alpha}} \left(\cos((\alpha - 1)\pi y + \eta_{\alpha,\beta_2})\right)^{\frac{1-\alpha}{\alpha}}, \tag{2.8}$$

where the constants η_{α,β_2} is defined as follows

$$\eta_{\alpha,\beta_2} := \beta_2 \frac{\pi}{2} K(\alpha), \tag{2.9}$$

and $K(\alpha)$ as in (2.5). Then the following theorem provides a direct way of generating $\mathcal{S}_\alpha(\sigma_2, \beta_2, \mu)$ RVs, see Weron [1996b] and Weron [1996a].

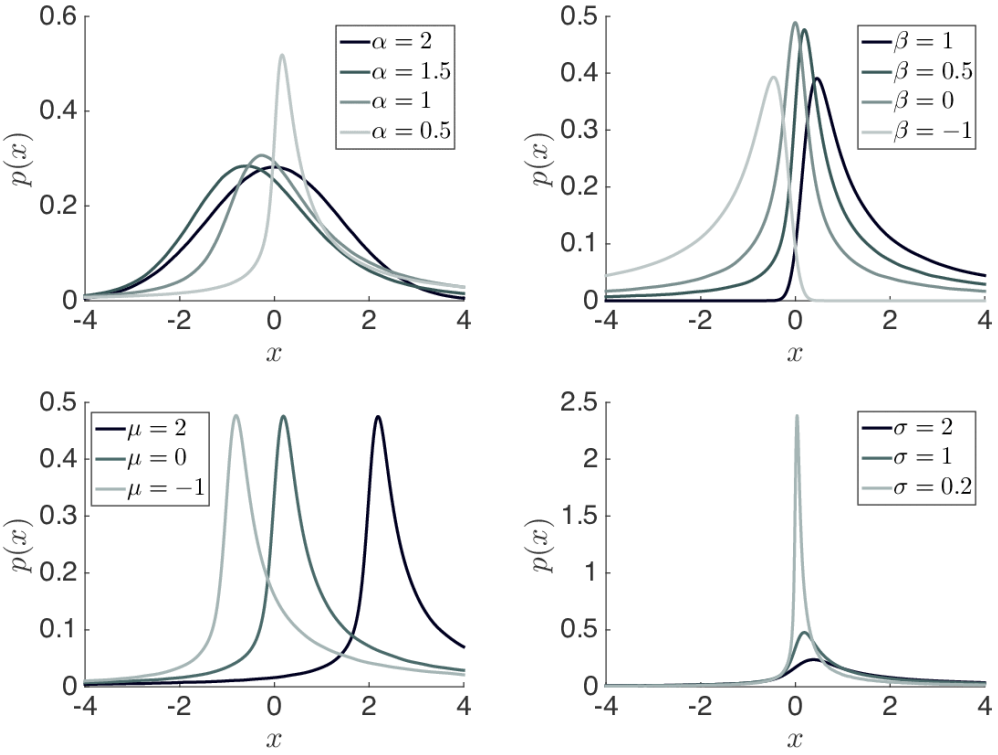


Figure 2.1 PDFs of the α-stable distribution $\mathcal{S}_\alpha(\sigma, \beta, \mu)$ for various parameter values. Whenever unspecified, $\alpha = 0.5$, $\sigma = 1$, $\beta = 0.5$, and $\mu = 0$.

Theorem 2.1.2. Let $\alpha \in (0, 2]$ and $\beta_2 \in [-1, 1]$ and let $t_{\alpha,\beta_2}(y)$ be defined as in (2.8). Let $Y \sim \mathcal{U}(-1/2, 1/2)$ and $W \sim \mathcal{E}(1)$, independent of Y. Let

$$Z_2 := \begin{cases} t_{\alpha,\beta_2}(Y)\, W^{\frac{\alpha-1}{\alpha}}, & \alpha \neq 1, \\ \left(\frac{\pi}{2} + \beta_2 \pi Y\right) \tan(\pi Y) - \beta_2 \log\left(\frac{W \cos(\pi Y)}{\frac{\pi}{2} + \beta_2 \pi Y}\right), & \alpha = 1. \end{cases}$$

Then $Z_2 \sim \mathcal{S}_\alpha(1, \beta_2, 0)$, with CF given by (2.4).

Remark 3. (Non-standard stable variables) The variables obtained according to Theorems 2.1.1 and 2.1.2 have standard stable distribution ($\sigma = 1$ or $\sigma_2 = 1$ and $\mu = 0$). We can generate any stable random variable X by scaling Z with a non-negative constant and shifting the distribution by adding the location (see [Samoradnitsky and Taqqu, 1994, p. 11]):

$$X_1 := \sigma Z_1 + \mu \sim \mathcal{S}_\alpha(\sigma, \beta, \mu)$$
$$X_2 := \sigma_2 Z_2 + \mu \sim \mathcal{S}_\alpha(\sigma_2, \beta_2, \mu)$$

Figure 2.1 shows, for example, $\mathcal{S}_\alpha(\sigma, \beta, \mu)$ PDFs for different combinations of the four parameters of the distribution, produced through kernel smoothing histograms of 10^6 samples generated using the CMS method in Theorem 2.1.1.

2.2 Multivariate Stable Distributions

The first results on multivariate stable distributions are historically due to Feldheim [1937]; here we refer to the review and collection of results of [Samoradnitsky and Taqqu, 1994, Chapter 2]. The definition of stability in $\mathbb{R}^d$ is analogous to that in $\mathbb{R}^1$, so that all linear combinations of elements of a stable vector are univariate stable RVs, and all linear combinations of stable vectors are also stable vectors. However, the multidimensional stable distribution has a more complicated parametrization than the scalar case. In fact, the skewness and the scale parameters are replaced by the *spectral measure*, a finite measure function defined on the unit sphere in $\mathbb{R}^d$, $\mathcal{S}_d = \{\mathbf{s} \in \mathbb{R}^d : \|\mathbf{s}\| = 1\}$,

$$\xi : \mathcal{S}_d \to (0, \infty).$$

The multivariate stable distribution is also parametrized by a tail parameter α and a location parameter $\boldsymbol{\mu}$, that have a meaning similar to the scalar case (when we consider unidimensional projections of the vector). We then write $\mathbf{X} \sim \mathcal{S}_{\alpha,d}\{\boldsymbol{\mu}, \xi\}$ to denote that the d-dimensional vector $\mathbf{X}$ has multivariate stable distribution, with CF $\phi_{\mathbf{X}}(\mathbf{s})$, such that[5]

$$\log \phi_{\mathbf{X}}(\mathbf{s}) = -\int_{\mathcal{S}_d} |\mathbf{s}'\mathbf{t}|^\alpha \left\{1 - i \operatorname{sgn}(\mathbf{s}'\mathbf{t}) \tan \frac{\pi\alpha}{2}\right\} \xi(\mathrm{d}\mathbf{t}) + i\mathbf{s}'\boldsymbol{\mu}, \qquad \alpha \neq 1. \tag{2.10}$$

Notice that this is, again, only one of the possible parametrizations, corresponding to the scalar CF (2.3), which is recovered by taking $\xi : \{-1, 1\} \to (0, \infty)$ and

$$\sigma = (\xi(1) + \xi(-1))^{1/\alpha}, \qquad \beta = \frac{\xi(1) - \xi(-1)}{\xi(1) + \xi(-1)}.$$

A more recent review of the properties of multivariate stable distributions can be found in Nolan [2008]. For example, given a set $S \subset \mathcal{S}_d$, the mass that ξ assigns to S determines the tail behaviour of a stable vector in "direction" S (namely the mass assigned by the multivariate stable law to the cone generated by S). In contrast, the behaviour around the mode is determined by the integrand in (2.10).

[5] In our notation, $\mathbf{s}$ is a column vector and $\mathbf{s}'$ is its transpose.

2.2.1 Stable Vectors with Discrete Spectral Measure

Unlike the scalar case, it is generally not possible to exactly generate α-stable random vectors, except for few cases. One of these is the case of discrete spectral measures, of the form

$$\xi(\mathbf{s}) = \sum_{j=1}^{n} \gamma_j \delta_{\mathbf{s}_j}(\mathbf{s}), \tag{2.11}$$

where $\gamma_j > 0$ are weights, and $\delta_{\mathbf{s}_j}(\cdot)$ are point masses at the points $\mathbf{s}_j \in \mathcal{S}_d, j = 1, \ldots, n$. We refer to Nolan [2008] for the method for simulating a random vector with discrete spectral measure.

This class is extremely important, because it is dense in the class of all α-stable multivariate distributions. In fact, a multivariate stable distribution with arbitrary spectral measure can be approximated, up to a desired tolerance, by a distribution with appropriate discrete spectral measure, and Byczkowski et al. [1993] provide indications on how to choose the number, locations, and weights of the point masses for this purpose. As mentioned in Section 1.3, this is one of the reasons why the spectral measure is frequently assumed to be discrete in inference procedures for stable vectors: this reduces the dimensionality of the inference problem, while it guarantees closeness to the "true" distribution.

Contour plots of bivariate stable PDFs with $\alpha = 0.9$ and $\alpha = 1.6$, discrete spectral measure and $\boldsymbol{\mu} = 0$ are shown in Figure 2.2. The point masses are indicated with red dots in the plot, with relative size indicative of the weights; see Table 2.1 for the exact locations and weights values. The PDFs have been generated using the Matlab functions `mvstablediscspecmeas.m`, `mvstablepdf.m` and `mvstablernd.m` from the STABLE 5.3, C Library Version, software, available from Nolan [2018b].[6] In the case of discrete spectral measure, the PDF presents creases along the rays that connect the origin with the point masses; moreover, the PDF becomes 'more uniform' (less 'star-like') as $\alpha \to 2$ (the multivariate Gaussian distribution is recovered form the stable for $\alpha = 2$). For the reason mentioned in the previous section, we expect to see more stable points in the tails of the distribution, when we look at the cone generated by the point masses. In contrast, the behaviour around the mode is determined by the integrand in (2.10). The mode is at the origin when the point masses are symmetrically placed, while it is shifted otherwise.

A property useful for our purposes connects univariate and multivariate stable distributions with discrete spectral measure (see [Samoradnitsky and Taqqu, 1994, Example 2.3.6]): if $\mathbf{X} = \mathbf{AY}$, where the components of $\mathbf{Y}$ are independent α-stable

[6] Granted to us for research purposes.

RVs with the same tail parameter α, but not necessarily identically distributed, $Y_k \sim \mathcal{S}_\alpha(\sigma_k, \beta_k, \mu_k)$, and $\mathbf{A}$ is a matrix with appropriate dimensions, then $\mathbf{X}$ is an α-stable vector with discrete spectral measure concentrated on symmetric pairs of points of the unit sphere.[7] Hence, a vector formed by stacking independent α-stable random variables is an α-stable vector with discrete and symmetric spectral measure ($\mathbf{A}$ being the identity matrix).

[7] The vice-versa holds too, making this a characterization of symmetric discrete spectral measures, see [Samoradnitsky and Taqqu, 1994, p. 70].

Table 2.1 Summary of the parameters (point masses locations $\mathbf{s}_j$ and weights γ_j) of the PDFs of bivariate stable distributions in Figure 2.2.

Case	Parameters						
1	$\mathbf{s}_j$	(1,0)	$(1/2, \sqrt{3}/2)$	$(-1/2, \sqrt{3}/2)$	(-1,0)	$(-1/2, -\sqrt{3}/2)$	$(1/2, -\sqrt{3}/2)$
	γ_j	0.25	0.125	0.25	0.25	0.125	0.25
2	$\mathbf{s}_j$	(1,0)	$(\sqrt{3}/2, 1/2)$	$(1/2, \sqrt{3}/2)$	(0,1)	$(-1/2, \sqrt{3}/2)$	–
	γ_j	0.1	0.3	0.1	0.3	0.1	–

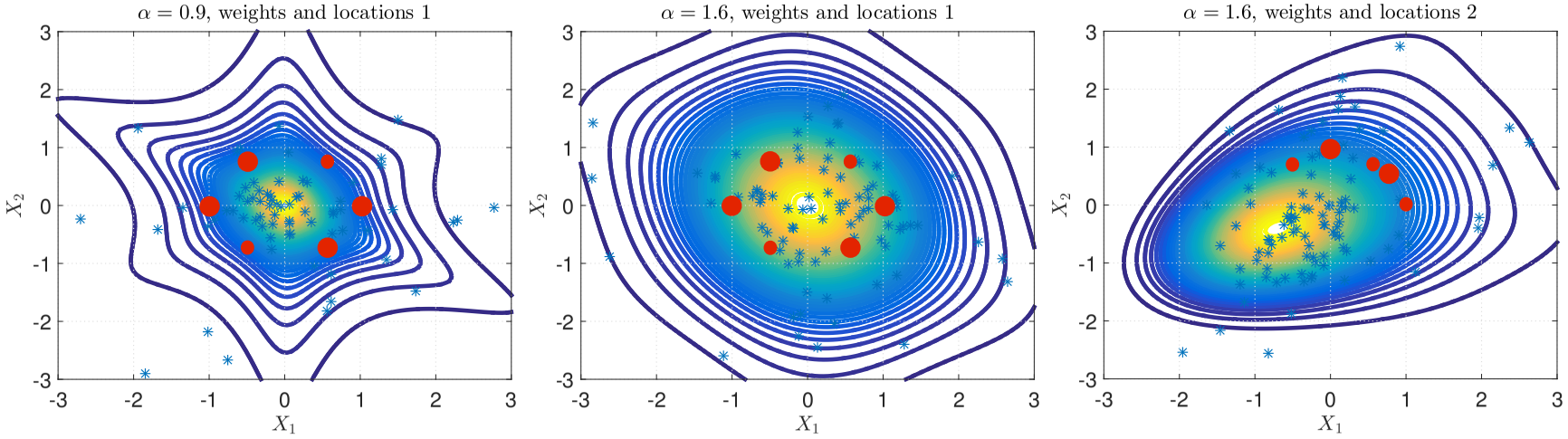


Figure 2.2 Contour plots of PDFs of a bivariate stable distribution with discrete spectral measures. We consider two values of the tail parameter ($\alpha = 0.9$ on the left, $\alpha = 1.6$ in the centre and on the right), and two configurations for the weighted point masses, as in Table 2.1). The blue stars are 100 samples generated from the distribution, while the red dots represent the weighted point masses (the relative size of each dot is indicative of its weight).

2.3 Stable Stochastic Processes

In this section we refer to [Samoradnitsky and Taqqu, 1994, Ch. 3], [Embrechts et al., 1997, Ch. 8.8] and Janicki and Weron [1994]. A stochastic process is a (possibly uncountable) sequence of RVs, $\{X(t)\}$, $t \in T$, where T is an arbitrary set, e.g. $\mathbb{R}$, $\mathbb{R}^n$, a set of sets or a set of functions.[8] A stochastic process is defined by specifying its *finite dimensional distributions* (FDD), i.e. the distributions of the vectors

$$(X(t_1), X(t_2) \ldots, X(t_d)), \qquad t_1, t_2, \ldots t_d \in T,\ d \geq 1,$$

and by proving their consistency see e.g. [Billingsley, 2008, Ch. 7, Sect. 36]. Then a stochastic process is stable iff all its FDDs are (multivariate) stable. The special case $\alpha = 2$ corresponds to Gaussian processes, uniquely identified by their mean and covariance functions.

2.3.1 Stable Lévy Processes

When $T = \mathbb{R}^+$, a class of great relevance is formed by Lévy processes (or motions), widely treated in the monographs of Bertoin [1998], Sato [1999] and Applebaum [2004], and in Barndorff-Nielsen et al. [2001] and Tankov and Cont [2015]. A Lévy process, denoted $\ell(t)$ in the following, is such that

(i) $\ell(0) = 0$;

(ii) $\{\ell(t)\}$ has independent increments: $\ell(t_2) - \ell(t_1), \ldots, \ell(t_d) - \ell(t_{d-1})$ are independent for any choice of of $0 \leq t_1 \leq \ldots \leq t_d$, and $d \geq 1$;

(iii) the increments are identically distributed: $\ell(t) - \ell(s) \stackrel{\mathcal{D}}{=} \ell(t-s)$, $t > s$.

In particular, we focus on stable Lévy processes, meaning that

$$\ell(t) - \ell(s) \sim \mathcal{S}_\alpha((t-s)^{1/\alpha}, \beta, 0), \tag{2.12}$$

for some $\alpha \in (0, 2]$ and $\beta \in [-1, 1]$. When $\alpha = 2$, $\ell(t)$ is the widely studied Brownian motion (see e.g. Borodin and Salminen [2012]; Hida [1980]; Revuz and Yor [2013]). Brownian motion is the only Lévy process with continuous trajectories, while stable Lévy processes[9] with $\alpha < 2$ are *pure jump* processes. However, neither Brownian motion nor other stable Lévy motions are differentiable.

[8] To simplify the notation, we do not indicate the dependence of the stochastic process on events in the underlying probability space. However, notice that we should formally write $X(t, \omega)$, $t \in T$, $\omega \in \Omega$ for the ensemble of the stochastic process.

[9] Also referred to as *infinite activity* of Lévy processes, see Tankov and Cont [2015].

A very desirable property for inference and simulation is $1/\alpha$-self-similarity, possessed by the class of stable Lévy motions[10], $\alpha \in (0, 2]$, meaning that the FDDs satisfy

$$(\ell(at_1), \ell(at_2), \ldots, \ell(at_d)) \stackrel{\mathcal{D}}{=} a^{1/\alpha} (\ell(t_1), \ell(t_2), \ldots, \ell(t_d)) , \tag{2.13}$$

for any choice of the values $t_1, t_2, \ldots t_d \in T$, $d \geq 1$, and $a > 0$. Quoting [Embrechts et al., 1997, p. 542], (2.13) means that, "if we interpret t as time and X_t as space, then every change of time scale $a > 0$ corresponds to a change of space scale $a^{1/\alpha}$. The smaller α, the more dramatic the change of space coordinate". Self-similarity corresponds to the fact that, in (2.12), the only parameter that is different for different time scales is the scale parameter[11] and it varies through a deterministic function of α, while the tail and skeweness parameter are a constant characteristic of the process. Such a property makes α-stable Lévy processes a very natural first approach towards generalizing the classical Brownian motion framework to the heavy-tailed case.

Observe, however, that the self-similarity of α-stable motions is not inherited by more general stable processes (including Guassian processes), such as those that are the solution of continuous time linear models driven by stable Lévy noise, and detailed in Sections 2.3.2 and 2.3.3. Nonetheless, quoting again [Embrechts et al., 1997, Example 8.9.3], "the underlying $1/\alpha$-self similarity of the Lévy motion creates patterns for the resulting process on a large or small scale", making it a similarly natural way for describing processes that represents the sum of many terms evolving e.g. through time.

Simulation of Stable Lévy Processes

We refer to [Tankov and Cont, 2015, Chapter 6] and references therein for an overview of simulation of Lévy processes, including the stable case.

A first way to approximately simulate a stable Lévy process is by time discretization, see also [Janicki and Weron, 1994, Ch. 6]. We define a grid of indices $t_1, \ldots, t_d \in (0, T]$, simulate the increments as in (2.12) by means of the CMS algorithm,[12] and sum them up. The distribution of the increments at different time scales is known, and these can be further simulated, if needed. Figure 2.3 shows some examples obtained by changing the stability index α. Observe that smaller values of α correspond to larger jumps in the trajectories. Observe also that, when $\alpha < 2$,

[10] This derives from the fact that every self-similar process can be obtained as a weak limit of normalized and centred i.i.d. sum processes, and α-stable Lévy motions are in the limit of a functional CLT, see e.g. [Embrechts et al., 1997, pp. 95, 545].

[11] A similar change affects the location, when this is non-zero.

[12] These are discrete-time realizations of the *noise process* $\{\mathrm{d}\ell(t)\}$

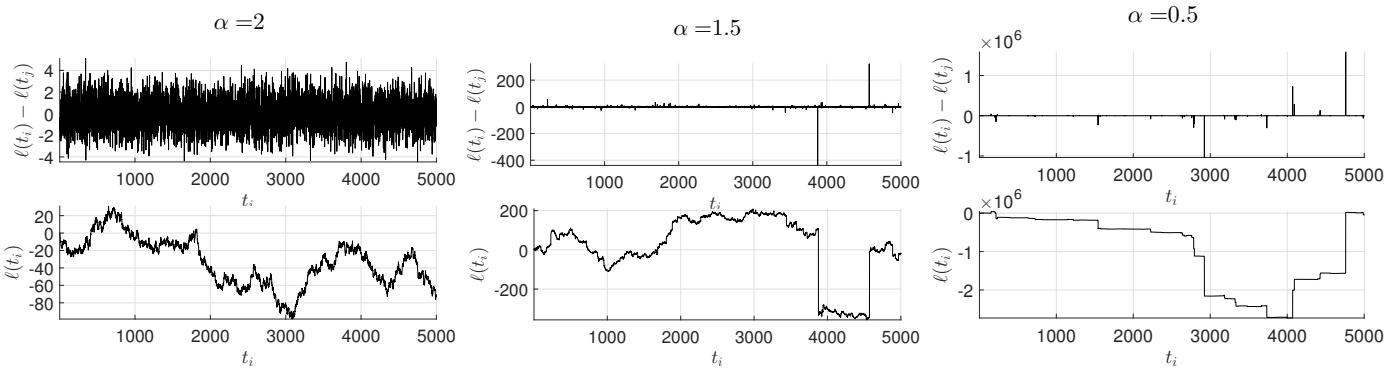


Figure 2.3 Realizations of discretized symmetric and standard α-stable noise (top - here $t_j = t_{i-1}$) and Lévy processes (bottom), varying α. The case $\alpha = 2$ corresponds to Brownian motion, while decreasing α gives larger jumps in the trajectory (notice the different ordinate scale in the three cases).

and especially for small values of α, even if the trajectories appear to be piece-wise constant, this is not the case, given that these processes have infinitely many small jumps on $[0, T]$.

A second way of approximately simulating a stable Lévy process, proposed by Asmussen and Rosiński [2001], is by approximating the large jumps through a compound Poisson process. Jumps smaller than some level $\epsilon > 0$ are truncated, and approximated by a Brownian motion term.

We do not investigate this simulation method, while we mention a third one, that derives from Poisson series representation (PSR) of stable Lévy processes, and was introduced by Rosiński [2001]. Truncations of the PSR correspond to truncating jumps smaller in magnitude than $c^{-1/\alpha}$, for $c > 0$, and this approach appears to have better convergence properties than the second one, when no compensation for the approximation is performed in neither scheme. Our main contribution in Chapter 5 is the approximation of the remainder term of the series for the PSR for stable RVs. In Chapter 9, we point out how such results extend directly to the (modified)[13] PSR for Lévy processes, so that accounting for the series residual further improves the convergence of this third approximate simulation method.

Moreover, in Chapter 8 we work on approximating the (modified) PSR of the solution to linear stochastic differential equations (SDEs) driven by scalar stable Lévy processes. Such an approach is different than the one above, in that not the driving noise, but directly the stable process resulting from linear dynamics is approximated, and we will explain how this facilitates inference methods. We start by giving a description of such models and processes in the following.

[13]The meaning of this notation is clarified in Chapter 8.

2.3.2 Continuous-Time Linear Models

In our developments, we are particularly interested in continuous-time extensions of the linear model (1.1), namely in the continuous-time autoregressive model of order P, CAR(P). These belong to the wider class of continuous-time linear models, satisfying the SDE

$$\mathrm{d}\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t)\,\mathrm{d}t + \mathbf{h}\,\mathrm{d}\ell(t), \tag{2.14}$$

where $\mathbf{x}(t) = [x_1(t), \ldots, x_P(t)]'$ is the, potentially vector-valued, state, $\mathbf{A}$ is a $P \times P$ matrix describing the interaction of the components of $\mathbf{x}(t)$, and $\mathbf{h}$ is a P-dimensional vector describing the effects of the stable noise process $\{\mathrm{d}\ell(t)\}$. The choice of the parameters $\mathbf{A}$ and $\mathbf{h}$ determines the class of the process, including CAR(P) processes, see Brockwell [2001, 2004]; Brockwell and Lindner [2009], as well as mean reverting processes and Langevin dynamics, see Lemons and Gythiel [1997]. We detail these cases in Chapter 8, together with the stability conditions for the corresponding deterministic dynamical system.

Continuous-time models are more challenging in terms of inference than the discrete time models (1.1), but they are also more realistic for scenarios where data are sampled at irregular time intervals. Notice however that the process is necessarily observed at discrete times $\{t_i\}$, and we consider having linear observations, in analogy with (1.3),

$$y(t_i) = \mathbf{b}'\mathbf{x}(t_i) + v(t_i), \tag{2.15}$$

where $\mathbf{b}$ is a P-dimensional vector and $\{v(t_i)\}$ is the observation noise process, which we assume to be Gaussian, $v(t_i) \sim \mathcal{N}(0, \sigma_v^2)$.

If $\{\ell(t)\}$ is a Brownian motion then $\{\mathbf{x}(t)\}$ is the Gaussian Ornstein–Uhlenbeck (OU) process introduced by Uhlenbeck and Ornstein [1930]. A wide range of results have been developed in the literature for OU processes, see e.g. Harvey [1990]; Øksendal [2003]. However, as for the discrete-time case, if the driving noise process is Gaussian, such models cannot account for heavy tails. These can be modelled by choosing $\{\ell(t)\}$ to be a non-Gaussian stable Lévy process. In this case, $\{\mathbf{x}(t)\}$ is a non-Gaussian OU process, see Barndorff-Nielsen et al. [2001], Barndorff-Nielsen and Shephard [2001], and Masuda [2004] for the vector formulation; see also Amini and Unser [2014]; Unser et al. [2014a,b] for more recent work on the comparison of Gaussian and non-Gaussian OU processes from an engineering perspective.

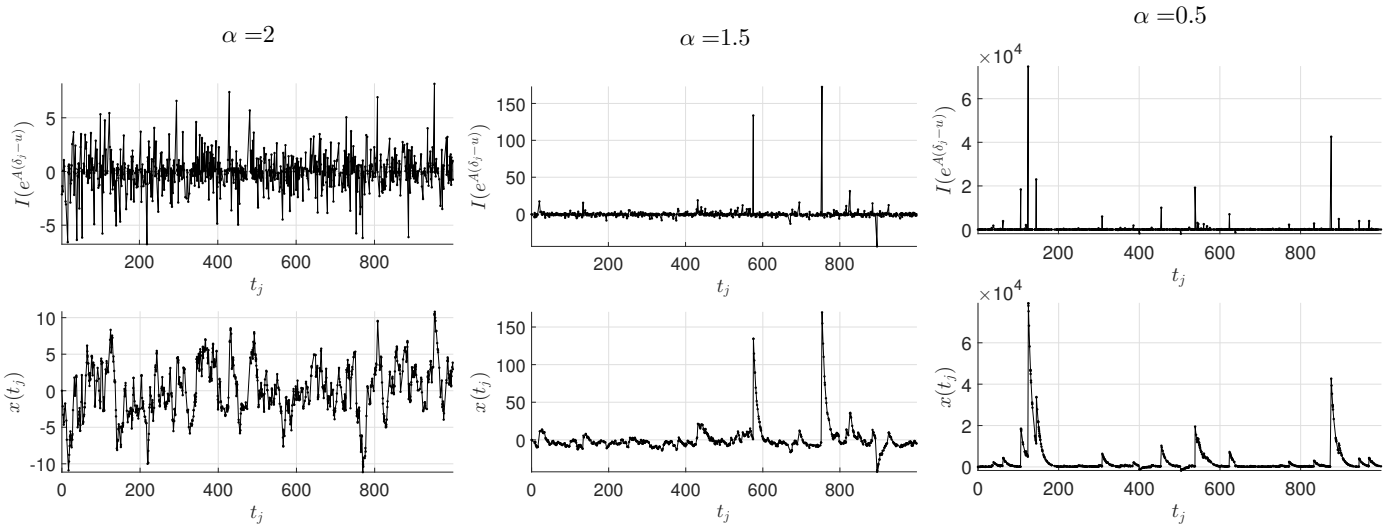


Figure 2.4 Realizations of discretized scalar α-stable integrals (top) and OU-processes (bottom), with $A = -0.1$, $\beta(u) = \beta = 0.8$, varying $\alpha \in \{2, 1.5, 0.5\}$.

The linear SDE (2.14) has solution

$$\mathbf{x}(t+\delta) = e^{\mathbf{A}\delta}\mathbf{x}(t) + \int_0^\delta e^{\mathbf{A}(\delta-u)}\mathbf{h}\,\mathrm{d}\ell(u), \tag{2.16}$$

see e.g. Kloeden and Platen [1992]. In the vector case ($P > 1$), the first term in the right-hand side of (2.16) contains $e^{\mathbf{A}\delta}$, a matrix exponential, defined by means of its Taylor series expansion. We provide considerations on the efficient computation of this term in Chapter 8. The second term in the right-hand side of (2.16) is a stable stochastic integral.

2.3.3 Stable Stochastic Integrals

Stable stochastic integrals have form

$$\mathbf{I}(\mathbf{f}) = \int_E \mathbf{f}(u)\,\mathrm{d}\ell(u). \tag{2.17}$$

This is the stable stochastic integral on the set E, of the deterministic, potentially vector valued, function $\mathbf{f} : \mathbb{R} \to \mathbb{R}^d$, namely a stochastic process indexed by functions, with stable FDDs, see [Samoradnitsky and Taqqu, 1994, Ch. 3.1]. We will work with the specific set of functions involved in the linear model (2.16), $\mathbf{f}(u) = e^{\mathbf{A}(\delta-u)}\mathbf{h}$, with $d = P$ and the integration domain $E = (0, \delta)$.

For the developments in Chapter 8, it is useful to recall that a stochastic integral can also be interpreted as integral against a random measure (RM), $M(\cdot)$. Random measures are, in turn, stochastic processes indexed by sets, see [Samoradnitsky and

Taqqu, 1994, Ch. 3.3]). To define a stable RM over a sample space Ω, we need to specify a control measure $m(\mathrm{d}u)$, such that Ω is m-finite, and a skeweness intensity function $\beta : \Omega \to [-1, 1]$. Then a stable Lévy motion is equivalent to the stable RM measure of the set $[0, t]$, $\ell(t) = M([0, t])$, with control measure the Lebesgue measure and constant skewness intensity $\beta(u) = \beta$. Hence $\mathbf{I}(\mathbf{f})$ in (2.17) can be interpreted as integral against the RM $M(\cdot)$ corresponding to the stable Lévy motion $\{\ell(t)\}$, so that we can write[14]

$$\mathbf{I}(\mathbf{f}) = \int_E \mathbf{f}(u) M(\mathrm{d}u). \tag{2.18}$$

Observe that, for the stable stochastic integral (2.17) to be well defined, it is necessary that

$$\int_E ||\mathbf{f}(u)||^\alpha m(\mathrm{d}u) < \infty, \qquad \alpha \in (0, 1) \cup (1, 2), \tag{2.19}$$

while we refer to [Samoradnitsky and Taqqu, 1994, Ch. 3.4] for the condition that holds when $\alpha = 1$.

Simulation of Stable Stochastic Integrals

If the integrand in (2.17) is fixed and scalar valued, $f : \mathbb{R} \to \mathbb{R}$, the distribution of its stochastic integral $I(f)$ is univariate α-stable, with the same tail exponent as that of the Lévy process $\{\ell(t)\}$ and the other parameters depending on α, f, and E, as in [Samoradnitsky and Taqqu, 1994, p. 117, 122]. The top row of Figure 2.4 shows realizations of the scalar process $I(e^{A(\delta_j - u)})$, with $A = -0.1$, $\delta_j = t_j - t_{j-1}$, and $0 = t_0 < t_1 < \ldots < t_N < T$ a discretized, uniformly sampled time grid ($T = N = 10^3$ in the figure); the values of the stochastic integral are sampled using the CMS method for generating scalar RVs, with accordingly set parameters. The corresponding (discretized) realization of the process $x(t_j)$, $x(0) = 0$ and $h = 1$ is displayed at the bottom, and this is obtained by adding the memory term $e^{A\delta} x(t)$, according to (2.16) (this is a simple exponential in the scalar case). We compare three different values of α, with smaller values of α causing larger jumps in the trajectories. In all the cases the process $\{x(t)\}$ reverts to its zero mean,[15] with spikes in correspondence of the spikes of the stochastic integral, but more smooth transitions thanks to the memory term.

[14] From this it also follows that $M([0, t]) = \int_0^T \mathbb{1}\left(x \in [0, t]\right) M(\mathrm{d}x) = I\left(\mathbb{1}[0, t]\right)$, $0 \le t \le T$.

[15] As expected considering the ordinary differential equation corresponding to (2.14), without the stochastic term, namely fixing $h = 0$.

Similarly, for a fixed vector-valued function $\mathbf{f} = [f_1, \ldots, f_d]'$, the stochastic integral (2.17) is the element-by-element integral

$$\mathbf{I}(\mathbf{f}) = \left[\int_E f_1(u)\,\mathrm{d}\ell(u),\ \int_E f_2(u)\,\mathrm{d}\ell(u),\ \ldots,\ \int_E f_d(u)\,\mathrm{d}\ell(u)\right]'$$

whose components have the same index of stability α, given that they are integrated against the same RM. Then, according to [Samoradnitsky and Taqqu, 1994, Proposition 3.4.2, p. 125], for a fixed $\mathbf{f}$, $\mathbf{I}(\mathbf{f})$ is a stable random vector, with location vector parameter $\boldsymbol{\mu} = \mathbf{0}$, and spectral measure depending on α, $\mathbf{f}$ and E, as in [Samoradnitsky and Taqqu, 1994, pp. 114-117]. Such a characterization of the spectral measure is not straightforward, and we do not report its expressions, because, unless it is discrete, it is not possible to exactly simulate from the process.[16] In Chapter 8 we provide an approach for approximately simulating (2.17), based on the PSR, that is valid regardless of the nature of its corresponding spectral measure. As anticipated in the introductory Section 1.3, besides allowing for simulations, our approach provides also a different parametrization to map the inference of an infinite-dimensional function to a finite dimensional estimation problem, alternative to its approximation via a weighted sum of point masses as in (2.11).

2.4 Latent Variable Representations for Stable Distributions

We now summarize the LV representations for the stable distribution that are used throughout the thesis. Our focus here is mainly on representations for scalar RVs, while we provide extensions to the vector case or to stochastic processes as needed.

[16] For example, following the consideration in Section 2.2.1, the spectral measure of $\mathbf{I}(\mathbf{f})$ is discrete and symmetric only if its elements are independent. According to [Samoradnitsky and Taqqu, 1994, p. 128-129], this is equivalent to requiring

$$\begin{cases} f_{k_1}(u) f_{k_2}(u) = 0 \quad m-\text{a.e.}, & \alpha \in (0,2), \\ \int_E f_{k_1}(u) f_{k_2}(u) = 0, & \alpha = 2, \end{cases}$$

for any subset $\{k_1, k_2\}$ of $\{1, \ldots, d\}$, namely that all the components of the vector-valued function $\mathbf{f}$ have disjoint support on E, a condition not generally met.

2.4.1 CMS and Marginal Representation of the PDF

The first LV model that we consider derives from the CMS method for generating stable RVs. We work with the $\mathcal{S}_\alpha(\sigma_2, \beta_2, 0)$ parametrization, because this was chosen by Buckle [1995], who first made this LV model explicit.[17]

Consider again Theorem 2.1.2, and the random variables $W \sim \mathcal{E}(1)$, $Y \sim \mathcal{U}(-1/2, 1/2)$ and $Z_2 \sim \mathcal{S}_\alpha(1, \beta_2, 0)$: the CMS method can be thought of as a deterministic transformation $T(W, Y) = (Z_2, Y)$, where

$$T : (0, \infty) \times (-1/2, 1/2) \longrightarrow (-\infty,\, 0) \times \left[-1/2,\, l_{\alpha,\beta_2}\right] \cup (0,\, \infty) \times \left[l_{\alpha,\beta_2},\, 1/2\right],$$

and

$$l_{\alpha,\beta_2} := -\eta_{\alpha,\beta_2}/\pi\alpha, \tag{2.20}$$

with η_{α,β_2} defined in (2.9). Notice that, while the variables W and Y are independent, Z_2 depends on Y through T. The joint PDF $p_{(Z_2,Y)}(z, y)$, can be obtained in closed form from $p_{(W,Y)}(w, y)$ by inverting T and computing the Jacobian of T^{-1}. If we make explicit also the dependence on α and β_2,[18] this leads to

$$p_{(Z_2,Y)} : (-\infty,\, 0) \times \left[-1/2,\, l_{\alpha,\beta_2}\right] \cup (0,\, \infty) \times \left[l_{\alpha,\beta_2},\, 1/2\right] \longrightarrow (0,\, \infty),$$

with

$$p_{(Z_2,Y)}(z,\, y|\alpha,\, \beta_2) = \frac{\alpha}{|\alpha - 1|} \exp\left\{-\left|\frac{z}{t_{\alpha,\beta_2}(y)}\right|^{\alpha/(\alpha-1)}\right\} \left|\frac{z}{t_{\alpha,\beta_2}(y)}\right|^{\alpha/(\alpha-1)} \frac{1}{|z|}, \tag{2.21}$$

and t_{α,β_2} defined in (2.8). The proof can be found in Buckle [1994] and Buckle [1995]. This is a state augmentation procedure:[19] it allows one to obtain closed-form expressions for the PDF of the augmented state by treating Y as an auxiliary variable, and the stable PDF of interest, $p_{Z_2}(z)$, can conceptually be obtained by

[17] A conceptually analogous integral representation is provided by Nolan [1997] to suite the $\mathcal{S}^0$ parametrization, based on the original formulation by Zolotarev [1966]. Numerical schemes for the computations of the integral (2.23) in both the $\mathcal{S}^0$ and $\mathcal{S}^1$ parametrizations are implemented in the STABLE 5.3, C Library Version, software, available from Nolan [2018b], and have been used for approximate frequentist inference.

[18] We stress that, in this representation, Z_2 is a standardized stable RV, hence $\sigma_2 = 1$ and $\mu = 0$. Stable RVs with arbitrary location and scale are obtained by simple linear transformation as in Remark 3.

[19] In the following the concepts of auxiliary or latent variable models and models with augmented state are used as synonyms.

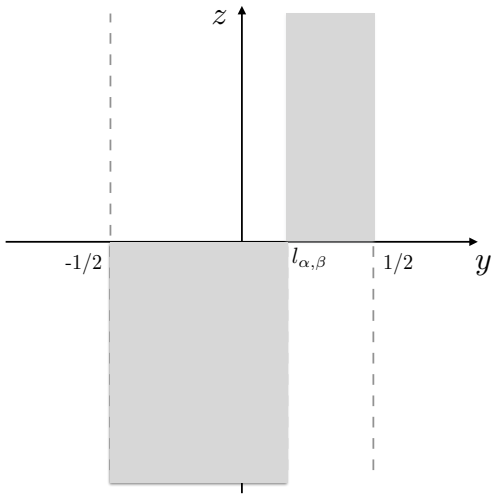


Figure 2.5 Illustration of the domain of $p_{(Z,Y)}$, the PDF of the vector (Z, Y), whose marginal component Z is a standard stable RV. We refer to (2.20) for the definition of $l_{\alpha,\beta}$.

marginalization. In fact, denoting with

$$\mathcal{Y} = \begin{cases} (l_{\alpha,\beta_2},\, 1/2), & Z_2 > 0, \\ (-1/2,\, l_{\alpha,\beta_2}), & Z_2 < 0, \end{cases} \tag{2.22}$$

the support of Y in the random vector (Z_2, Y), illustrated by the shaded area in Figure 2.5, it follows that the stable PDF has marginal representation[20]

$$p_{Z_2}(z) = \int_{\mathcal{Y}} p_{(Z_2,Y)}(z, y)\, \mathrm{d}y \tag{2.23}$$

$$= \int_{\mathcal{Y}} p_{(Z_2|Y)}(z|y) p_Y(y)\, \mathrm{d}y, \tag{2.24}$$

where in the second identity we use the marginal-conditional decomposition to split the joint PDF $p_{(Z_2,Y)}(z, y)$ into the product of the conditional $p_{(Z_2|Y)}(z|y)$ and the marginal $p_Y(y)$ PDFs. Observe that, due to the high degree of non-linearity of (2.21), the integral (2.23) does not have an analytic solution. However, the fact the the integrand $p_{(Z_2,Y)}(z, y)$ can be evaluated is a key factor for simulation based inference methods, as shown in Chapters 3 and 4.

[20] We exclude the case $Z_2 = 0$, but this is a point with zero Lebesgue measure, and the stable PDF in this point can be obtained by continuity.

2.4.2 Product Property and Scale Mixture of Normals (SMiN)

A different LV model derives from the product property for stable RVs, see for example [Feller, 1966, p. 176] or [Samoradnitsky and Taqqu, 1994, p. 20-21]. We summarize the product property in the following Lemma.

Lemma 1. Let X_1 and $X_2 > 0$ be strictly stable RVs, with tail parameters α_1 and α_2, respectively.[21] Then the product $X = X_1 X_2^{1/\alpha_1}$ has stable distribution with tail parameter $\alpha_1\alpha_2$.

Lemma 1 holds for both symmetric and asymmetric stable distributed X_1, and it is possible to define all the parameters of X in terms of those of X_1 and X_2. However, its most practical use for inference purposes is when X_1 is symmetric. In this case, it follows that X is also symmetric, meaning that it is possible to transform any symmetric stable distribution with stability index α_1 to another symmetric stable distribution with stability index $\alpha_1\alpha_2 < \alpha_1$. In particular, recalling that the Gaussian distribution is a special case of symmetric stable distribution, according to the product property, it is possible to transform any Gaussian RV

$$X_1 \sim \mathcal{S}_2(\sigma, 0, 0) \overset{\mathcal{D}}{=} \mathcal{N}(0, 2\sigma^2),$$

into $X \sim \mathcal{S}_\alpha(\sigma, 0, 0)$, by choosing

$$X_2 \sim \mathcal{S}_{\alpha/2}((\cos(\pi\alpha/4))^{2/\alpha}, 1, 0), \tag{2.25}$$

and $X = X_1\sqrt{X_2}$.[22]

The consequence of the product property is that symmetric stable RVs can be thought of as a continuous *scale mixture of normals*, SMiN, a concept that was originally introduced by Andrews and Mallows [1974]. A RV X is a SMiN if we can write the following LV model

$$X|S \sim \mathcal{N}(0, \tilde{\sigma}^2 S), \tag{2.26a}$$

$$S > 0 \sim \mathcal{D}_S, \tag{2.26b}$$

[21] Observe that the requirements on X_2 imply that X_2 is a fully right skewed stable RV, with $\alpha_2 < 1$.

[22] Nonzero location parameters can be handled separately, by simple shift of the SMiN representation, namely considering $X + \mu$. Here we consider $\mu = 0$, as this is the assumption in the inference schemes presented in Chapter 3.

where S is the (positive) mixing RV, with distribution $\mathcal{D}_S$, and $\tilde{\sigma} > 0$ is a constant.[23] Equivalently, we can write that $X = \sqrt{S}T$, where $T \sim \mathcal{N}(0, \tilde{\sigma}^2)$. If we assume that $\mathcal{D}_S$ admits a continuous PDF $p_S(s)$, then the PDF of X can be obtained by marginalizing the joint PDF $p_{(X,S)}(x, s) = p_{(X|S)}(x|s)p_S(s)$ with respect to S, where $p_{(X|S)}(x|s)$ is a Gaussian PDF, as in (2.26a), namely[24]

$$p_X(x) = \int_{\mathbb{R}^+} \mathcal{N}(0, \tilde{\sigma}^2 s) p_S(s)\, \mathrm{d}s. \tag{2.27}$$

The product property that transforms a Gaussian into a *symmetric* stable RV is then equivalent to a SMiN representation with $T = X_1$, $\tilde{\sigma} = \sqrt{2}\sigma$ and $S = X_2$. Observe that the SMiN model is also a state augmentation procedure. The main difference with the marginal representation (2.23) stemming from the CMS is that, not only the joint PDF of the stable variable and the auxiliary variable $p_{(X,S)}(x, s)$ has closed form, but it is possible to generate the LV S exactly and in a simple way. Recall, in fact, that S is a fully right skewed stable RV, and that it can be sampled through the CMS algorithm in Section 2.1.1. Thus the SMiN representation enables simulation based inference schemes similar to, but potentially more efficient than, those allowed by representation (2.23), as shown in the Chapter 3.

Sub-Gaussian Random Vectors

A similar conditionally Gaussian representation also holds for the class of α-stable sub-Gaussian random vectors and stochastic processes, for which we refer to [Samoradnitsky and Taqqu, 1994, p. 77, 142]. if S is of the form (2.25), and $\mathbf{T} = [T_1, \ldots, T_d]'$ is a d-dimensional, zero-mean, Gaussian random vector independent of S, and with variance-covariance matrix $\mathbf{\Sigma}$, then $\mathbf{X} = \boldsymbol{\mu} + \sqrt{S}\mathbf{T}$ is a sub-Gaussian symmetric stable random vector, with tail parameter α, location parameter $\boldsymbol{\mu}$,[25] and spectral measure defined as in [Samoradnitsky and Taqqu, 1994, p.82]. The PDF of a sub-Gaussian stable random vector has marginal expression similar to (2.27)[26]

$$\begin{aligned} p_{\mathbf{X}}(\mathbf{x}) &= \int_{\mathbb{R}^+} p_{(\mathbf{X}|S)}(\mathbf{x}|s) p_S(s)\, \mathrm{d}s \\ &= \int_{\mathbb{R}^+} \mathcal{N}(\boldsymbol{\mu}, s\mathbf{\Sigma}) p_S(s)\, \mathrm{d}s. \end{aligned} \tag{2.28}$$

[23] For example, the Student's t-distribution with ν degrees of freedom is a SMiN with $\tilde{\sigma} = 1$, and $S \sim \mathrm{IG}(\nu/2, \nu/2)$, the inverse gamma distribution with shape and scale parameters equal to $\nu/2$.

[24] With an abuse of notation, here we indicate $\mathcal{N}(\mu, \sigma^2)$ the PDF of a Gaussian RV with mean μ, and variance σ^2.

[25] We consider nonzero location here, because this is the assumption in the inference schemes in the literature that we present in Chapter 3.

[26] With another abuse of notation, here we indicate with $\mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$ the PDF of a Gaussian RV with mean vector $\boldsymbol{\mu}$, and variance-covariance matrix $\mathbf{\Sigma}$.

As mentioned in Section 1.3, the hypothesis of sub-Gaussianity simplifies the structure of the distribution parameters of $\mathbf{X}$: besides α and $\boldsymbol{\mu}$, instead of inferring the spectral measure function, one needs to infer only the matrix $\boldsymbol{\Sigma}$ (the parameters of S are either constant or a function of α, as in (2.25)). Observe, however, that, while all symmetric stable RVs have SMiN representation, this is not the case for symmetric vectors (and stochastic processes), only a subclass of which has sub-Gaussian structure. Hence we recall some Bayesian inference methods developed for this class, but we do not make separate analysis on this class in our developments.

2.4.3 Poisson Series Representation and Mean and Scale Mixture of Normals (MaSMiN)

Unfortunately, no conditionally Gaussian representation arises from the product property for *skewed* stable RVs. However, it is possible to cast a different representation, the Poisson series representation (PSR) of α-stable RVs (and processes), into a *mean and scale mixture of normals*, MaSMiN, model, now valid also for the skewed stable distribution.

Poisson Series Representation

The PSR was originally introduced by Lévy and formalised by LePage et al. [1981] and LePage [1981, 1989], and for this reason is also known as LePage series representation. Here we refer to [Samoradnitsky and Taqqu, 1994, Theorem 1.4.5 on p. 28]. In the parametrization (2.3), an α-stable RV $X \sim \mathcal{S}_\alpha(\sigma, \beta, 0)$, $\alpha \in (0, 2)$, admits the following infinite random summation representation

$$X \overset{\mathcal{D}}{=} \sum_{j=1}^{\infty} \Gamma_j^{-1/\alpha} W_j - \mathbb{E}[W_1] b_j^{(\alpha)}, \tag{2.29}$$

where $\mathbb{E}[\cdot]$ denotes the expected value operator and where

- $\{\Gamma_j\}_{j=1}^{\infty}$ are the arrival times of a unit rate Poisson process, so that the differences $\Gamma_j - \Gamma_{j-1}$, are i.i.d. Exponential RVs with mean 1:

$$\Gamma_j - \Gamma_{j-1} \overset{i.i.d.}{\sim} \mathcal{E}(1), \qquad j = 1, 2, \ldots \tag{2.30}$$

  with the convention that $\Gamma_0 := 0$;

- $\left\{W_j\right\}_{j=1}^{\infty}$ are i.i.d. RVs , mutually independent of $\{\Gamma_j\}_{j=1}^{\infty}$, with

$$\mathbb{E}[|W_1|^\alpha] < \infty; \tag{2.31}$$

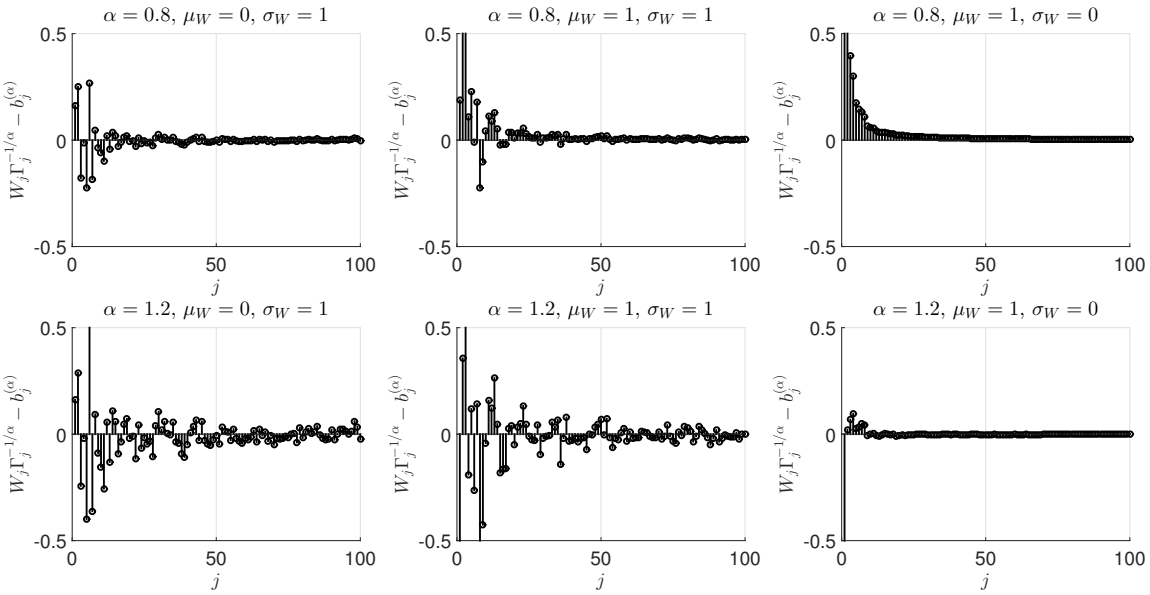


Figure 2.6 First 100 terms of a PSR realization, with $W_1 \sim \mathcal{N}(\mu_W, \sigma_W^2)$ in three different scenarios for μ_W and σ_W^2, summarized in Table 2.2. Top row: $\alpha = 0.8$; bottom row: $\alpha = 1.2$, corresponding to a less heavy-tailed distribution than $\alpha = 0.8$.

- $\{b_j^{(\alpha)}\}_{j=1}^{\infty}$ are deterministic centering coefficients, non-zero only if $\alpha \in [1, 2)$:

$$b_j^{(\alpha)} = \begin{cases} \dfrac{\alpha}{\alpha - 1}\left(j^{\frac{\alpha-1}{\alpha}} - (j-1)^{\frac{\alpha-1}{\alpha}}\right), & \alpha \in (1, 2) \\ \mathbb{E}\left[W_1 \int_{W_1/j}^{W_1/(j-1)} x^2 \sin x \, \mathrm{d}x\right] / \mathbb{E}[W_1], & \alpha = 1. \end{cases} \tag{2.32}$$

  The telescoping structure of the coefficients $\{b_j^{(\alpha)}\}$ leads to the following expression for their sum, when $\alpha \in (1, 2)$

$$\sum_{j=1}^{N} b_j^{(\alpha)} = \frac{\alpha}{\alpha - 1} N^{\frac{\alpha-1}{\alpha}}, \quad \forall N \geq 1. \tag{2.33}$$

Observe that (2.29) is a series representation for *strictly stable* RVs (the location parameter of X is $\mu = 0$), and a stable RV with non-zero location parameter μ can be obtained as $X + \mu$. We also stress that (2.29) is not valid when $\alpha = 2$, corresponding to the Gaussian distribution. Moreover, in the following we do not consider the case $\alpha = 1$, for which simple, but different, expressions of the quantities of interest hold.

Reparametrization Induced by the PSR.

The PSR induces a mapping between the α-th absolute moments of W_1 and the parameters $\{\sigma, \beta\}$ of the stable distribution, as reported in [Samoradnitsky and

Taqqu, 1994, p. 28]

$$\sigma^\alpha = \frac{\mathbb{E}\left[|W_1|^\alpha\right]}{C_\alpha}, \tag{2.34a}$$

$$\beta = \frac{\mathbb{E}\left[|W_1|^\alpha \operatorname{sgn} W_1\right]}{\mathbb{E}\left[|W_1|^\alpha\right]}, \tag{2.34b}$$

where

$$C_\alpha = \left(\int_0^\infty x^{-\alpha} \sin x \, \mathrm{d}x\right)^{-1} = \frac{1-\alpha}{\Gamma(2-\alpha)\cos(\pi\alpha/2)}, \tag{2.35}$$

and $\Gamma(\cdot)$ is the Gamma function, defined in (B.12), Appendix B.4. We refer to Appendix A.1 for considerations on the transformations (2.34a) and (2.34b), with focus on the case $W_1 \sim \mathcal{N}(\mu_W, \sigma_W^2)$, the non-central normal distribution with mean μ_W and variance σ_W^2, which will be the case of interest in the rest of the thesis. In such a case, there is a mapping

$$\{\alpha, \mu_W, \sigma_W, \mu\} \longmapsto \{\alpha, \sigma, \beta, \mu\},$$

given that the moments of W_1 are uniquely determined by μ_W and σ_W^2. Specifically, the central Gaussian distribution ($\mu_W = 0$) corresponds to the symmetric stable distribution ($\beta = 0$), while the degenerate Gaussian distribution ($\sigma_W = 0$ and either $\mu_W > 0$ or $\mu_W < 0$) corresponds to the fully skewed stable distribution (with either $\beta = 1$ or $\beta = -1$).

Figure 2.6 shows, for example, the first 100 terms $W_j \Gamma_j^{-1/\alpha} - b_j^{(\alpha)}$ of PSR realizations, with $W_j \sim \mathcal{N}(\mu_W, \sigma_W^2)$, in the three different scenarios for μ_W and σ_W^2 summarized in Table 2.2. We also compare two values of α: $\alpha = 0.8$ and $\alpha = 1.2$, corresponding to a more heavy tailed and less heavy tailed scenario, respectively (the same RVs Γ_j and W_j are used in the two cases). Notice that the summands in the PSR are stochastically decaying (in absolute value). In fact, when $\{W_j\}_{j=1}^\infty$ are degenerate RVs ($\sigma_W = 0$), the increasing nature of the sequence $\{\Gamma_j\}_{j=1}^\infty$ makes $\{\Gamma_j^{-1/\alpha}\}_{j=1}^\infty$ to be monotonically decreasing (the convergence is faster as α decreases). When the $\{W_j\}_{j=1}^\infty$ are non-degenerate, their magnitude affects the strict monotonicity of the summands. Table 2.2 also reports the numerical results of the transformations (2.34a) and (2.34b) for the sets of parameters used in Figure 2.6.

Table 2.2 Numerical mappings between the parameters of W_1, μ_W and σ_W, and the α-distribution parameters, σ and β, when $\alpha = 0.8$ and $\alpha = 1.2$, as in Figure 2.6.

	μ_W	σ_W	σ	β
$\alpha = 0.8$	0	1	1.16	0
	1	1	1.71	0.84
	1	0	1.42	1
$\alpha = 1.2$	0	1	1.37	0
	1	1	1.99	0.99
	1	0	1.80	1

MaSMiN Representation for Stable RVs

The reason why we focus on the case $W_1 \sim \mathcal{N}(\mu_W, \sigma_W^2)$, is that this allows a conditionally Gaussian representation of the α-stable distribution which is useful for inference as detailed in Chapter 3. In fact, from the PSR (2.29) it follows that, if $W_1 \sim \mathcal{N}(\mu_W, \sigma_W^2)$, we can write an auxiliary variables model for X as

$$\begin{aligned} X|\{\Gamma_j\}_{j=1}^{\infty} &\sim \mathcal{N}\left(\mu_W M, \sigma_W^2 S^2\right), \\ M &:= \sum_{j=1}^{\infty} \Gamma_j^{-1/\alpha} - b_j^{(\alpha)}, \qquad (2.36) \\ S^2 &:= \sum_{j=1}^{\infty} \Gamma_j^{-2/\alpha}. \qquad (2.37) \end{aligned}$$

In this model M and S^2 are treated as two *dependent* auxiliary RVs, with structure determined by the PSR, and X is conditionally Gaussian. Observe that the LVs in this model are formed by the infinite sequence $\{\Gamma_j\}$ or, equivalently, by the two RVs M and S^2. Thus, an alternative to the above model is

$$\begin{aligned} X|(M,S) &\sim \mathcal{N}\left(\mu_W M, \sigma_W^2 S^2\right), \qquad (2.38a) \\ (M, S > 0) &\sim \mathcal{D}_{(M,S)}, \qquad (2.38b) \end{aligned}$$

where M and S^2 are defined as in (2.36) and (2.37), and $\mathcal{D}_{(M,S)}$ is the joint distribution of (M, S), that we assume has PDF $p_{(M,S)}(m, s)$. Then the stable PDF has marginal representation

$$p_X(x) = \int_{\mathbb{R}^+} \int_{\mathbb{R}} \mathcal{N}\left(\mu_W m, \sigma_W^2 s^2\right) p_{(M,S)}(m, s)\, \mathrm{d}m\, \mathrm{d}s, \qquad (2.39)$$

in analogy with (2.27) deriving from the SMiN model.

While the exact representation of the stable law (2.39) is a very appealing framework in theory, it is computationally intractable. In fact, the joint distribution $\mathcal{D}_{(M,S)}$ cannot explicitly be characterized, and, more importantly, it is not possible to exactly sample M and S, because of the infinite summations involved in (2.36) and (2.37). This means that approximations of the MaSMiN model are needed, if this is chosen as the representation of a stable RV for simulation or inference purposes.

2.4.4 Approximations of the MaSMiN Model for Stable RVs

Three approaches have been presented in the literature for approximating the PSR of stable RVs. All of these are based on simulating only a finite number of the variables $\{\Gamma_j\}$ involved in (2.29), (2.36) and (2.37), driven by the fact that the addends of the PSR are stochastically decaying.

Truncation of the PSR

A first intuitive approach is to truncate[27] the PSR. Specifically, only the summands corresponding to the variables

$$\mathbf{\Gamma}_{(0,c)} := \{\Gamma_j\}_{j:\Gamma_j<c} \equiv \{\Gamma_j < c\} \equiv \{\Gamma_j\}_{j=1}^{N_{(0,c)}} \tag{2.40}$$

for $c > 0$, are simulated and summed up. Here c is a threshold chosen in such a way that the truncated PSR is 'close enough' to X, while $N_{(0,c)}$ is a random integer that makes the identity in (2.40) hold, namely the number of Poisson arrivals before time c.[28]

The truncation of the PSR induces an approximation of the series (2.36) and (2.37) with[29]

$$M_{(0,c)} := \sum_{j=1}^{N_{(0,c)}} \Gamma_j^{-1/\alpha}, \tag{2.41}$$

$$S^2_{(0,c)} := \sum_{j=1}^{N_{(0,c)}} \Gamma_j^{-2/\alpha}. \tag{2.42}$$

[27] This is also related to the third method for approximate generation of Lévy processes mentioned in Section 2.3.1, see [Tankov and Cont, 2015, Chapter 6].

[28] In Chapter 5 we show that $N_{(0,c)}$ has Poisson distribution.

[29] The subscript notation might seem redundant at first, but we use consistently with following Chapters.

This approach leads, in turn, to the approximate conditionally Gaussian model for $X \sim \mathcal{S}_\alpha(\sigma, \beta, 0)$

$$X|(M_{(0,c)}, S_{(0,c)}) \overset{\text{approx}}{\sim} \mathcal{N}\left(M_{(0,c)}, S^2_{(0,c)}\right), \tag{2.43a}$$

$$(M_{(0,c)}, S_{(0,c)} > 0) \sim \mathcal{D}_{(M_{(0,c)}, S_{(0,c)})}, \tag{2.43b}$$

where $X \overset{\text{approx}}{\sim} \mathcal{D}$ indicates that the (conditional in our case) distribution of the RV X is approximated by $\mathcal{D}$. Alternatively, we say that X is (conditionally on the realizations of $M_{(0,c)}$ and $S_{(0,c)}$) approximated by $X_{(0,c)} \sim \mathcal{N}\left(m_{(0,c)}, s^2_{(0,c)}\right)$. Then the PDF of X can be approximately represented as

$$p_X(x) \approx \int_{\mathbb{R}^+} \int_{\mathbb{R}} \mathcal{N}\left(m_{(0,c)}, s^2_{(0,c)}\right) p_{(M_{(0,c)}, S_{(0,c)})}(m_{(0,c)}, s_{(0,c)}) \, \mathrm{d}m_{(0,c)} \, \mathrm{d}s_{(0,c)},$$

where $p_X(x) \approx g(x)$ denotes that the PDF p_X is approximated by the function g.

This conditionally Gaussian structure of the truncated PSR is a favourable result for inference, in that it resembles the exact MaSMiN (2.39), and it can be used in practice, because the LVs associated with the truncated model are $\mathbf{\Gamma}_{(0,c)}$, or, equivalently, $M_{(0,c)}$ and ${S^2}_{(0,c)}$, and can be simulated. However, we anticipate that its convergence to the true PSR is 'slow', in that a large number of summands ($N_{(0,c)} \gg 1$) has to be generated for the truncated PSR $X_{(0,c)}$ to be close enough to X for any given value of the tail parameter α. A number of works in the literature has been devoted to the study the convergence of this approximation, see e.g. Ledoux and Paulauskas [1996] or Janicki and Weron [1994] and Janicki and Kokoszka [1992], and we will recall the main results in Chapter 7.

Better results can be obtained by (probabilistically) accounting for the unsampled summands in the PSR corresponding to the variables $\{\Gamma_j\}_{j=(N_{(0,c)}+1)}^{\infty}$. Two possible solutions have been proposed in Lemke [2014]; Lemke and Godsill [2011, 2012, 2014]; Lemke et al. [2015] for this purpose,[30] that maintain a conditionally Gaussian structure for the stemming approximation of the PSR.

Gaussian Approximation Approach

The first scheme is named the 'Gaussian approximation approach' (GAA). In this scheme, the residual of the PSR after $N_{(0,c)}$ terms, denoted $R_{(c,\infty)}$, is approximated with $\hat{R}_{(c,\infty)}$, an accordingly chosen Gaussian RV. The role of $\hat{R}_{(c,\infty)}$ is thus to compensate for the unsampled terms directly in the PSR. We denote with $\hat{X}_1$ the approximation of the PSR corresponding to the first $N_{(0,c)}$ terms of (2.29), to which

[30] An alternative method is given in Azzaoui and Clavier [2010], aimed at forward simulation of stable law variables and processes, but not, to our knowledge, at inference tasks.

the approximate residual $\hat{R}_{(c,\infty)}$ is added. A precise definition of these quantities is provided in Chapter 5, given that this approximation scheme is central to many analyses of this thesis.

The LVs involved in the GAA scheme are thus $\mathbf{\Gamma}_{(0,c)}$, defined as in (2.40), and $\hat{R}_{(c,\infty)}$, and they allow to define the moments $(\hat{M}_1, \hat{S}_1^2)$ of a conditionally Gaussian model for the stable RV $\hat{X}_1$. It will be clear from Chapter 5 that the RV $\hat{R}_{(c,\infty)}$ does not need to be sampled in order to compute $(\hat{M}_1, \hat{S}_1^2)$, because only its moments are needed, and these can be computed analytically as a function of c, making it possible to write

$$X|\mathbf{\Gamma}_{(0,c)} \overset{\text{approx}}{\sim} \mathcal{N}\left(\hat{M}_1, \hat{S}_1^2\right). \tag{2.44a}$$

Gaussian Approximation of the Moments Approach

A second scheme aims at approximating the series that define M (2.36) and S^2 (2.37) in the conditionally Gaussian representation of a stable RV X. This scheme is named the 'Gaussian approximation of the moments approach' (GAMA). In particular, if we denote with $\mathbf{R}_{(c,\infty)}$ the bivariate vector of residuals of these two series after $N_{(0,c)}$ terms, then the GAMA scheme approximates $\mathbf{R}_{(c,\infty)}$ with $\hat{\mathbf{R}}_{(c,\infty)}$, an, again appropriately chosen, bivariate Gaussian random vector. We denote with $\hat{M}_2$ and $\hat{S}_2^2$ the RVs formed by adding the elements of the random vector $\hat{\mathbf{R}}_{(c,\infty)}$ to the the first $N_{(0,c)}$ terms of M and S^2. Then $\hat{X}_2$ is the approximation of the PSR that is conditionally Gaussian with moments $\hat{M}_2$ and $\hat{S}_2^2$.

Unlike the GAA, in this approximate model the LVs are given by $\mathbf{\Gamma}_{(0,c)}$, defined as in (2.40), and $\hat{\mathbf{R}}_{(c,\infty)}$, that now needs to be sampled. It is then possible to write

$$X|\mathbf{\Gamma}_{(0,c)}, \hat{\mathbf{R}}_{(c,\infty)} \overset{\text{approx}}{\sim} \mathcal{N}\left(\hat{M}_2, \hat{S}_2^2\right). \tag{2.45a}$$

We will provide more details on the GAMA scheme in Chapter 8, where we work on the PSR for multivariate linear SDEs. Here we observe that, in the scalar case, one of the advantages over the GAA scheme is that the GAMA is able to capture the dependence structure of M and S through the extra-diagonal terms in the covariance matrix of $\hat{\mathbf{R}}_{(c,\infty)}$. Moreover, the GAMA scheme allows for marginalization of the likelihood of the stable distribution wrt the parameters μ, μ_W and σ_W^2, and this is useful for Bayesian inference methods, as summarized in Chapter 3. However, the fact that the residual $\hat{\mathbf{R}}_{(c,\infty)}$ needs to be simulated to construct $\hat{M}_2$ and $\hat{S}_2^2$ does require some extra computational complexity wrt the GAA scheme, especially for

inference purposes. Moreover, the resulting RV $\hat{S}_2^2$ is not guaranteed to be positive,[31] while this is a necessary condition for a variance term.

Both of these schemes thus have relative advantages and disadvantages. We expand on the GAA scheme in Chapters 5, 6, 7, where we quantify the error incurred by such approximation, provide guidance for the selection of the truncation parameter c, and show that it is likely to be superior to simple truncation of the PSR. Our analyses will refer to the symmetric stable distribution, but have the potential to be extended to the generically skewed case, which is the primary purpose of using the PSR over the SMiN representation. Furthermore, we leave to future work a similar analysis for the GAMA scheme.

In both cases, the ability to generate approximations of M and S allows for exact inference schemes for the corresponding approximate RVs $\hat{X}_1$ and $\hat{X}_2$ (i.e. approximate schemes for X), as summarized in Section 3.3.4 of the following chapter.

[31] The probability that $\hat{S}_2^2$ takes a negative value is, however, very small in practice.

Chapter 3

Background II: Bayesian Parameter Inference for Stable Distributions

This chapter summarizes some general principles and methods of Bayesian inference and how these have been applied in the literature to the LV representations for the α-stable distribution.

3.1 Bayesian Parameter Inference

The estimation of the parameters $\boldsymbol{\theta} \in \Theta$ of a distribution is an important step for the successful application of any chosen model.[1] Our motivation for focussing on Bayesian inference procedures is that they allow for natural uncertainty quantification. In this section we refer mainly to Robert and Casella [2004].

Uncertainty quantification is of high relevance in a framework of extreme valued data. Consider, for example, the task of estimating the tail parameter α when the data are thought to follow an α-stable model. As mentioned in Chapter 1.3, there exist several frequentist approaches in the literature for this purpose. However, besides the errors introduced by approximations of the likelihood function, the estimates produced by any such method will be affected by the empirical value of the sample at hand, $\boldsymbol{x} = [x_1, \ldots, x_N]$, especially when the sample size N is small. This problem is mitigated when the Baysian approach to inference is adopted, by treating the parameters $\boldsymbol{\theta}$ as RVs. The *prior* guess of the distribution of the parameters,

[1] Given that we are following a model based approach, we would need to specify that all our inference procedures depend on the selected model $\mathcal{M}$ (α-stable in this case). However we omit to indicate such dependence, given that in this dissertation we do not tackle model selection problems.

$p(\boldsymbol{\theta})$,[2] is automatically updated to form the *posterior* distribution of the parameters, $p(\boldsymbol{\theta}|\boldsymbol{x})$, after observing the data and computing their likelihood $p(\boldsymbol{x}|\boldsymbol{\theta})$, via the Bayes' rule:

$$p(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{x})} \propto p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}) := \tilde{p}(\boldsymbol{\theta}|\boldsymbol{x}), \tag{3.1}$$

where

$$p(\boldsymbol{x}) := \int_{\Theta} p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} \tag{3.2}$$

is the marginal likelihood of the data, or model evidence, namely the likelihood of the given dataset for any possible value of the parameters in the chosen model (in fact the parameters are integrated out). This is the normalizing constant of the posterior distribution and can thus be omitted in the right-hand side term of (3.1), to give the unnormalized posterior $\tilde{p}(\boldsymbol{\theta}|\boldsymbol{x})$. Observe furthermore that, under the assumption of independent data, the likelihood is factorized

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{n=1}^{N} p(x_n|\boldsymbol{\theta}),$$

and, in a LV setting such as (2.24), (2.27) or (2.39),

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \int_{\mathbf{Y}} p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{y} = \prod_{n=1}^{N} \int_{\mathcal{Y}_n} p(x_n, y_n|\boldsymbol{\theta})\,\mathrm{d}y_n \tag{3.3}$$

or, alternatively,

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \int_{\mathbf{Y}} p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{y} = \prod_{n=1}^{N} \int_{\mathcal{Y}_n} p(x_n|y_n, \boldsymbol{\theta})p(y_n|\boldsymbol{\theta})\,\mathrm{d}y_n, \tag{3.4}$$

where $\boldsymbol{y} = [y_1, \ldots, y_N]$ is an N-dimensional vector of LVs, one per data point, and $\boldsymbol{y} \in \mathbf{Y} = \prod_{n=1}^{N} \mathcal{Y}_n$, where $\mathcal{Y}_n$ is the integration domain of each LV y_n. Observe that $\mathcal{Y}_n$ can vary as a function of z_n, as in (2.22); observe, furthermore, that each y_n is a potentially vector valued LV, as in the case of the MaSMiN model (2.39).

From the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{x})$, point estimates such as the posterior average

$$\hat{\boldsymbol{\theta}}_A := \mathbb{E}_{p(\boldsymbol{\theta}|\boldsymbol{x})}[\boldsymbol{\theta}] = \int_{\Theta} \boldsymbol{\theta} p(\boldsymbol{\theta}|\boldsymbol{x})\,\mathrm{d}\boldsymbol{\theta}, \tag{3.5}$$

[2] For simplicity of notation, in this section we assume that all the distributions that we deal with are continuous and admit density functions, and we can thus interchange the terms *distribution* and *density*.

where $\mathbb{E}[\cdot]$ denotes the expected value,[3] or the posterior mode(s)

$$\hat{\boldsymbol{\theta}}_M := \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmax}}\, p(\boldsymbol{\theta}|\boldsymbol{x}) \tag{3.6}$$

or the posterior median can be computed.[4] It is often not possible to compute such estimators in closed form because it is not possible to solve the integration and maximization problems involved in (3.1), (3.2), (3.5) and (3.6).

However, if a set of (i.i.d.) samples from the posterior

$$\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots \boldsymbol{\theta}_M \overset{i.i.d.}{\sim} p(\boldsymbol{\theta}|\boldsymbol{x})$$

is available, Monte Carlo approximations of $p(\boldsymbol{\theta}|\boldsymbol{x})$ can be computed as follows

$$p(\boldsymbol{\theta}|\boldsymbol{x}) \approx \hat{p}(\boldsymbol{\theta}|\boldsymbol{x}) := \frac{1}{M}\sum_{j=1}^{M} \delta_{\boldsymbol{\theta}_j}(\boldsymbol{\theta}),$$

where $\delta_a(x)$ is the Dirac delta function centred at a. The corresponding Monte Carlo approximation of expected value of any integrable function $h(\cdot)$ against the posterior is

$$\mathbb{E}_{p(\boldsymbol{\theta}|\boldsymbol{x})}[h(\boldsymbol{\theta})] \approx \mathbb{E}_{\hat{p}(\boldsymbol{\theta}|\boldsymbol{x})}[h(\boldsymbol{\theta})] = \frac{1}{M}\sum_{j=1}^{M} h(\boldsymbol{\theta}_j).$$

It is possible to prove that $\mathbb{E}_{\hat{p}(\boldsymbol{\theta}|\boldsymbol{x})}[h(\boldsymbol{\theta})]$ is an unbiased and strongly consistent[5] estimator of $\mathbb{E}_{p(\boldsymbol{\theta}|\boldsymbol{x})}[h(\boldsymbol{\theta})]$. Moreover, a CLT holds for $\mathbb{E}_{\hat{p}(h(\boldsymbol{\theta})|\boldsymbol{x})}[h(\boldsymbol{\theta})]$, namely

$$\sqrt{M}\left(\mathbb{E}_{\hat{p}(\boldsymbol{\theta}|\boldsymbol{x})}[h(\boldsymbol{\theta})] - \mathbb{E}_{p(\boldsymbol{\theta}|\boldsymbol{x})}[h(\boldsymbol{\theta})]\right) \underset{M\to\infty}{\xrightarrow{\mathcal{D}}} \mathcal{N}(0, \operatorname{Var}_{p(\boldsymbol{\theta}|\boldsymbol{x})}[h(\boldsymbol{\theta})]), \tag{3.7}$$

where $\underset{M\to\infty}{\xrightarrow{\mathcal{D}}}$ denotes convergence in distribution as the number of samples M grows to infinity, and $\operatorname{Var}[\cdot]$ denotes the variance. The main advantage of Monte Carlo schemes is that the rate of convergence in (3.7) holds regardless of the dimensionality of Θ, unlike the convergence of numerical integration schemes, that typically degrades when the dimensionality of the parameter space increases.

[3] Subscript notation indicates here the distribution with respect to which we are taking the expected value.

[4] Each of these estimator can be interpreted as the minimizer of a Bayesian loss functions, see Berger [2013] or Lehmann and Casella [2006] for a reference.

[5] Meaning that the strong law of large numbers (SLLN) applies to $\mathbb{E}_{\hat{p}(\boldsymbol{\theta}|\boldsymbol{x})}[h(\boldsymbol{\theta})]$.

3.1.1 Monte Carlo Methods

The main issue with Monte Carlo approximations is that, except for a few special cases (inverse transform, composition and other 'ad hoc' methods, see Devroye [1986]), it is not possible to generate i.i.d. samples from the 'target' posterior $p(\boldsymbol{\theta}|\boldsymbol{x})$ (3.1) and hence to compute the above estimators (3.5), (3.6). Two methods that cope with this problem are the rejection sampling, RS, and the importance sampling, IS, algorithms. Both methods require being able to sample from an auxiliary 'proposal' distribution, q, that we assume can be evaluated pointwise.[6]

Rejection Sampler

The RS scheme generates a sample X from a probability distribution with PDF $p(x)$, $x \in \mathcal{X}$ (the posterior distribution for Bayesian inference purposes), if there are another distribution with PDF $q(x)$ and a constant $K \geq 1$ such that

$$p(x) \leq Kq(x), \qquad \forall x \in \mathcal{X}. \tag{3.8}$$

The method proceeds by simulating two RVs, $X \sim q$ and $U|X = x \sim \mathcal{U}[0, Kq(x)]$, where $\mathcal{U}[0, Kq(x)]$ denotes the uniform continuous distribution on the interval $[0, Kq(x)]$, until $U < p(x)$.[7] Then X is marginally distributed according to the target p. Algorithm 1 provides a schematic summary; if a sample of size M needs to be produced, the scheme is repeated M times. It is possible to prove that the number of draws before accepting a proposed value X is geometrically distributed with success probability $1/K$, and, hence, an average of K draws from q are required before obtaining a sample from p. For 'difficult'[8] target distributions p, the constant K that guarantees that (3.8) holds uniformly is typically large, meaning a long running time before a sample is produced.

A technique proposed in the literature to increase the probability of success is that of adaptive rejection sampling (ARS). ARS was initially developed by Gilks and Wild [1992] for exact sampling from distributions with log-concave densities: starting from an initial proposal distribution, ARS builds an envelope proposal PDF composed of truncated exponential PDFs, each inside an element of a partition of the support of the target distribution. The rejected points are added to modify the

[6] Here we consider normalized target and proposal densities, but the rejection and importance sampling schemes can be devised for unnormalized densities as well. In particular, the RS method is not affected by lack of normalization, while a bias is introduced in the self-normalized IS. However, in this case this bias typically leads to lower variance, according to a common trade off.

[7] This is equivalent to generating $X \sim q$ and $U \sim \mathcal{U}[0, K]$, until $U < p(x)/q(x)$.

[8] For example highly spiked and multimodal PDFs p.

Algorithm 1 RS - Rejection sampler

1: Simulate $X \sim q$, $U|X = x \sim \mathcal{U}[0, Kq(x)]$
2: **if** $U \leq p(x)$ **then**
3: **return** X
4: **else**
5: return to 1.

partition and the proposal, increasing the chance of acceptance at the following RS trial.

Importance Sampling

If only the expectation of a function h with respect to a distribution p needs to be estimated, as in (3.5), the IS scheme can be used instead of the RS scheme. Let $\mathcal{X}$ be the support of p, $\tilde{\mathcal{X}}$ be the support of q, also called 'importance distribution', and let $w(x) := p(x)/q(x)$ for $x \in \mathcal{X}$ be the likelihood ratio, or 'weight function'. If $\mathcal{X} \subseteq \tilde{\mathcal{X}}$ the following identity holds

$$\begin{aligned}\mathbb{E}_p[h(X)] = \int_{\mathcal{X}} h(x)p(x)\,\mathrm{d}x &= \int_{\tilde{\mathcal{X}}} h(x)p(x)\,\mathrm{d}x \\ &= \int_{\tilde{\mathcal{X}}} h(x)\frac{p(x)}{q(x)}q(x)\,\mathrm{d}x = \mathbb{E}_q[h(X)w(X)].\end{aligned}$$

Then a Monte Carlo estimate of $\mathbb{E}_p[h(X)]$ is

$$\mathbb{E}_p[h(X)] \approx \mathbb{E}_{\hat{q}}[h(X)w(X)] := \frac{1}{M}\sum_{j=1}^{M} h(X_j)w(X_j), \tag{3.9}$$

where $X_1, \ldots X_M \overset{i.i.d.}{\sim} q$, and $\hat{q}$ is the sample approximation of q. It is straightforward to prove that $\mathbb{E}_{\hat{q}}[h(X)w(X)]$ is an unbiased estimator for $\mathbb{E}_p[h(X)]$, and that its variance is

$$\mathrm{Var}_q[\mathbb{E}_{\hat{q}}[h(X)w(X)]] = \frac{1}{M}\left[\int_{\tilde{\mathcal{X}}} \frac{(h(x)p(x))^2}{q(x)}\,\mathrm{d}x - \mathbb{E}_p[h(X)]^2\right]. \tag{3.10}$$

The last expression provides indications on how to choose a good importance function. Although finite variance is not necessary for the convergence of $\mathbb{E}_{\hat{q}}[h(X)w(X)]$ in a SLLN sense, the IS scheme performs poorly in practice when (3.10) is not bounded. In fact, in this case, the estimator (3.9) will change abruptly in different realizations, even when $M \gg 1$.

A sufficient condition that guarantees that the variance (3.10) is finite, regardless of the choice of the function h, is

$$\frac{p(x)}{q(x)} = w(x) < \infty, \qquad \forall x \in \mathcal{X}, \tag{3.11}$$

which is always satisfied in the scenario we assumed $\mathcal{X} \subseteq \tilde{\mathcal{X}}$. This means that the importance distribution needs to have heavier tails than the target distribution. It is furthermore possible to prove that, if we consider only the variance as optimality criteria, the optimal importance distribution has PDF $q(x) = |h(x)|p(x)/\mathbb{E}_p[|h(X)|]$; however, this is not a practical choice because it requires knowing $\mathbb{E}_p[h(X)]$, which is the integral of interest. A good q is then one that has heavier tails than p, and it is nearly proportional to $|h(x)|p(x)$, presenting spikes where $|h(x)|p(x)$ does.

Observe also that condition (3.11) is weaker than (3.8), so that, when the latter holds, it is possible to compute expectations using both the RS and the IS schemes. The difference is that, in the RS method, only a subset of the samples generated from q will be retained, namely the samples that are distributed exactly according to the target distribution p. On the contrary, all the samples will be used in the IS scheme, but they will need to be weighted. In this scenario, it is not clear which method provides a Monte Carlo estimator with lower variance, see [Robert and Casella, 2004, p. 103] for a discussion.

3.1.2 Markov Chain Monte Carlo Methods

When the support of the target distribution p is high dimensional, it becomes difficult to choose a proposal distribution q that gives low rejection rates in the RS or low variance of the IS estimator, making these schemes often impractical. A solution that works better in these scenarios is to construct a Markov chain that has p as its invariant distribution.

We consider discrete-time Markov chains, namely stochastic processes $X^{(k)}$ indexed by the natural numbers $k \in \mathbb{N}$, with values in $\mathcal{X}$, the support of p, and whose FDDs obey the Markov property. In particular, given an initial distribution for the value of the chain at time 0, $X^{(0)} \sim p_0$, the probability that the chain transits to the state $x^{(k)}$ at time k, for any $k \geq 1$, given the past trajectory, depends only on the current value of the chain at time $k-1$

$$\begin{aligned} p_{(k,k-1)} &:= \mathbb{P}(X^{(k)} = x^{(k)} | X^{(k-1)} = x^{(k-1)}, \dots, X^{(1)} = x^{(1)}, X^{(0)} = x^{(0)}) \\ &= \mathbb{P}(X^{(k)} = x^{(k)} | X^{(k-1)} = x^{(k-1)}). \end{aligned}$$

Algorithm 2 MH - Metropolis-Hastings

1: Initialize $X^{(0)}$
2: **for** $k = 1 : N_{it}$ **do**
3: given $X^{(k-1)}$, generate $X' \sim q(\cdot|X^{(k-1)})$
4: compute $\rho(X^{(k-1)}, X') := 1 \wedge \dfrac{\tilde{p}(X')q(X^{(k-1)}|X')}{\tilde{p}(X^{(k-1)})q(X'|X^{(k-1)})}$
5: take $X^{(k)} = \begin{cases} X' & \text{with probability} \quad \rho(X^{(k-1)}, X'), \\ X^{(k-1)} & \text{with probability} \quad 1 - \rho(X^{(k-1)}, X') \end{cases}$

Given such a transition kernel, the *stationary* or invariant distribution of a Markov chain is a probability distribution over the state space $\mathcal{X}$ that is preserved under transitions.[9]

We furthermore consider Markov chains that are *irreducible* (the chain can go from any state to any other), *aperiodic* (the chain is not forced in cycles of a finite number of steps between states) and *positive recurrent* (the expected hitting time for any state is finite, and hence the chain almost surely visits every state infinitely often). In fact, in this case, a SLLN holds for Monte Carlo estimators built through realizations of the Markov chains, and that aim at estimating expectations of functions w.r.t. the stationary distribution. The combination of Markov chains with Monte Carlo methods give rise to the homonymous set of methods (MCMC); we refer e.g. to Robert and Casella [2004] and Gilks et al. [1995b] for a review.

MCMC techniques result thus useful if it is possible to construct an irreducible Markov chain (namely an initial distribution p_0 and a Markov transition kernel $p_{(k,k-1)}$) with a given target p as its invariant distribution, where p (or its unnormalized version $\tilde{p}$) can be only pointwise evaluated, as in the case of the parameter posterior distribution.

Metropolis-Hastings Sampler

The most commonly used schemes that exploit the above principle are based on the construction of Metropolis-Hastings (MH) transition kernels, originally proposed in the physical sciences by Metropolis et al. [1953] and Hastings [1970]. Given a target distribution p, a MH transition kernel is defined by specifying a proposal distribution with density $q : \mathcal{X} \times \mathcal{X} \to \mathbb{R}^+$, potentially dependent on the current state. Given the current state of the Markov chain $X^{(k-1)}$, the MH algorithm samples a new value X'

[9] Hence, when the initial distribution of the chain coincides with the stationary distribution, the Markov chain is a stationary process.

from $q(\cdot|X^{(k-1)})$ and accepts it as $X^{(k)}$ with *acceptance probability*

$$\rho(X^{(k-1)}, X') := 1 \wedge \frac{\tilde{p}(X')q(X^{(k-1)}|X')}{\tilde{p}(X^{(k-1)})q(X'|X^{(k-1)})}, \tag{3.12}$$

where $x \wedge y$ denotes the minimum between x and y. If the proposed value is rejected, the chain maintains the value of $X^{(k-1)}$ at time k. This procedure is repeated for N_{it} iterations, as summarized in Algorithm 2. It is possible to prove that the Markov chain thus obtained is irreducible if the proposal density assigns non-zero mass to all sets that have non-zero mass under the target p. Furthermore the chain is invariant because it is *reversible*, a condition met, in turn, because MH transition kernels satisfy the *detailed balance equation*.

Some common proposal distributions, and the respective acceptance probabilities, are as follows:

- symmetric proposals $q(X'|X^{(k-1)}) = q(X^{(k-1)}|X')$, with $\rho(X^{(k-1)}, X') = 1 \wedge \frac{\tilde{p}(X')}{\tilde{p}(X^{(k-1)})}$;
- independent proposals $q(X'|X^{(k-1)}) = q(X^{(k-1)})$, with $\rho(X^{(k-1)}, X') = 1 \wedge \frac{\tilde{p}(X')q(X^{(k-1)})}{\tilde{p}(X^{(k-1)})q(X')}$;
- random walk proposals $q(X'|X^{(k-1)}) = q(X' - X^{(k-1)})$, that becomes $q(X'|X^{(k-1)}) = q(|X' - X^{(k-1)}|)$ in the symmetric case.

The proposal distribution is the key element of choice when applying the MH scheme. In fact, the proposal distribution should guarantee a 'good' mixing of the Markov chain, namely an exploration of the state space $\mathcal{X}$ that is not too 'local' (typically high acceptance rate but small steps) and not too 'global' (rejection of the proposed values for many iterations, and only seldom acceptance of a new value). This is quantified by the decay of the autocorrelation function of the chain (as detailed in Section 3.1.3), and by the fact that the chain is not 'trapped' in a local mode of the posterior. Moreover, the computational burden of the proposal mechanism determines the complexity of the sampler, making it a crucial choice for practical implementability of the method.

For these reasons, a vast literature is devoted to the design and study of properties of proposal distributions. We refer, for example, to Roberts et al. [2001] for the 'optimal' tuning of random walk proposals when the target is a multivariate normal distribution. Adaptive mechanisms, in which the proposal is updated along the

Algorithm 3 GS - Gibbs sampler

1: Initialize $(X_1^{(0)}, \dots, X_d^{(0)})$
2: **for** $k = 1 : N_{it}$ **do**
3: given $(X_1^{(k-1)}, \dots, X_d^{(k-1)})$, generate
4: $X_1^{(k)} \sim p_1(X_1 | X_2^{(k-1)}, \dots, X_d^{(k-1)})$
5: $X_2^{(k)} \sim p_2(X_2 | X_1^{(k)}, X_3^{(k-1)}, \dots, X_d^{(k-1)})$
6: $\vdots$
7: $X_d^{(k)} \sim p_d(X_d | X_1^{(k)}, \dots, X_{d-1}^{(k)})$

process using the full information cumulated,[10] have also been widely studied, see e.g. Haario et al. [2001] and Andrieu and Thoms [2008]. Other proposal designs extend the ARS mechanism described in Section 3.1.1 to non-log-concave targets, see Gilks et al. [1995a] and, more recently, Martino et al. [2012]. In this case the adaptively constructed proposal (now not an envelope on the target in the log-scale) is used as proposal in a MH accept-reject step. Observe moreover that, when geometric information about the target distribution is available (such as the gradient), using this information leads to typically improved samplers. This is the basic principle of Hamiltonian Monte Carlo methods, introduced by Duane et al. [1987]; see also Neal et al. [2011] and Betancourt [2017] for recent developments.

Gibbs Sampler

Different Markov transition kernels can be combined to obtain hybrid chains, for example through mixtures or cycles of kernels. One such case is the Gibbs sampling (GS) scheme, proposed by Geman and Geman [1984] and promoted by Gelfand and Smith [1990]. The Gibbs sampler can be viewed, in fact, as a cycle of d MH kernels, each whit acceptance probabilities equal to 1, where d is the dimension of the state space.[11] At each iteration k, the GS iteratively draws a set of d samples from the *full conditional* distributions

$$p_i(X_i^{(k)} | X_1^{(k)}, X_2^{(k)}, \dots, X_{i-1}^{(k)}, X_{i+1}^{(k-1)}, \dots, X_d^{(k-1)}), \qquad i = 1, \dots, d \tag{3.13}$$

[10] This breaks the Markov property. However, such proposals maintain the correct ergodic properties of the estimators, and are, for this reason, used in common practice.

[11] We assume sampling of one component of the state at the time, but strategies that group the variables in 'blocks' are widely studied.

Algorithm 4 MH-GS - Metropolis within Gibbs sampler

1: Initialize $(X_1^{(0)}, \dots, X_d^{(0)})$
2: **for** $k = 1 : N_{it}$ **do**
3: **for** $1 = 1 : d$ **do**
4: given $(X_1^{(k-1)}, \dots, X_d^{(k-1)})$
5: generate $\tilde{X}_i^{(k)} \sim q_i(\tilde{X} | X_1^{(k)}, X_2^{(k)}, \dots, X_{i-1}^{(k)}, X_i^{(k-1)}, X_{i+1}^{(k-1)}, \dots, X_d^{(k-1)})$
6: compute $\rho = 1 \wedge \left\{ \frac{\left(\frac{p_i(\tilde{X}_i^{(k)} | X_1^{(k)}, X_2^{(k)}, \dots, X_{i-1}^{(k)}, X_{i+1}^{(k-1)}, \dots, X_d^{(k-1)})}{q_i(\tilde{X}_i^{(k)} | X_1^{(k)}, X_2^{(k)}, \dots, X_{i-1}^{(k)}, X_i^{(k-1)}, X_{i+1}^{(k-1)}, \dots, X_d^{(k-1)})} \right)}{\left(\frac{p_i(X_i^{(k-1)} | X_1^{(k)}, X_2^{(k)}, \dots, X_{i-1}^{(k)}, X_{i+1}^{(k-1)}, \dots, X_d^{(k-1)})}{q_i(X_i^{n-1} | X_1^{(k)}, X_2^{(k)}, \dots, X_{i-1}^{(k)}, \tilde{X}_i^{(k)}, X_{i+1}^{(k-1)}, \dots, X_d^{(k-1)})} \right)} \right\}$
7: take $X_i^{(k)} = \begin{cases} \tilde{X}_i^{(k)} & \text{with probability} \quad \rho, \\ X_i^{(k-1)} & \text{with probability} \quad 1 - \rho \end{cases}$

where the values $(X_1^{(k)}, X_2^{(k)}, \dots, X_{i-1}^{(k)})$ are realizations of the respective RVs in the current iteration, while the values $(X_{i+1}^{(k-1)}, \dots, X_d^{(k-1)})$ are realizations in the previous iteration, as described in Algorithm 3.[12]

When drawing samples from the full conditional distributions is not possible, but their pointwise evaluation is available, targeting the full conditionals with a MH step provides a valid scheme, as proposed in Muller [1991] and Müller [1992]. Such a composition is called Metropolis-within-Gibbs sampler (MH-GS): instead of simulating $X_i^{(k)}$ from (3.13), we simulate $\tilde{X}_i^{(k)}$ from an auxiliary distribution $q_i(\tilde{X}_i^{(k)} | X_1^{(k)}, X_2^{(k)}, \dots, X_{i-1}^{(k)}, X_i^{(k-1)}, X_{i+1}^{(k-1)}, \dots, X_d^{(k-1)})$, $i = 1, \dots, d$ and perform an accept-reject step, with ratio similar to (3.12), as detailed in Algorithm 4.

3.1.3 MCMC Diagnostics

If a Markov chain $\{X_i\}_{i=1}^{N_{it}}$ converges to a stationary distribution p, but it is not initialized at it, then a number of iterations is required for the convergence to happen. These iterations are also known as *burn-in* time, and they should be discarded when applying the SLLN for estimating $\mathbb{E}_p[h(X)]$ through $\mathbb{E}_{\hat{p}}[h(X)]$, where $\hat{p}$ is the Monte Carlo estimate of p obtained from the chain

$$\hat{p}(x) := \frac{1}{N_{it}} \sum_{i=0}^{N_{it}} \delta_{X_i}(x). \tag{3.14}$$

[12] We assume deterministic and ordered scans for clarity of notation, but random scans are a valid alternative.

There are no systematic ways to determine when a Markov chain has reached convergence in distribution, and the burn-in is usually found through visual inspection of the traceplots of the chain.

We assume now that the chain has been initialized at the stationary distribution (or that burn-in iterations have been discarded). Under further (sufficient) conditions, a CLT applies to $\mathbb{E}_{\hat{p}}[h(X)]$, analogous to (3.7) for i.i.d samples. Such conditions relate to the rate of convergence of the transition probability of a chain to the stationary distribution (geometric or uniform ergodicity), and to either the existence of second or higher order moments of $h(X)$ under p, or to reversibility of the chain. We refer, for example, to [Robert and Casella, 2004, Ch. 6.6, 6.7] and further references therein. When the CLT applies,

$$\sqrt{N_{it}}\left(\mathbb{E}_{\hat{p}}[h(X)] - \mathbb{E}_{p}[h(X)]\right) \underset{N_{it}\to\infty}{\xrightarrow{\mathcal{D}}} \mathcal{N}\left(0, \sigma_h^2\right),$$

where the limiting variance

$$\sigma_h^2 := \mathrm{Var}_p[h(X_0)] + 2\sum_{i=1}^{\infty} \mathrm{Cov}_p(h(X_0), h(X_i)),$$

has now an additional term with respect to that in (3.7), caused by the correlation existing between variables. The quantity

$$\tau_h := \frac{\sigma_h^2}{\mathrm{Var}_p[h(X_0)]} = 1 + 2\sum_{i=1}^{\infty} \mathrm{Corr}_p(h(X_0), h(X_i)),$$

where Corr$[\cdot]$ denotes the correlation, is known as *integrated auto-correlation time* (IACT) for estimating $\mathbb{E}_p[h(X)]$, using the given Markov chain $\{X_i\}$. It has the interpretation that a Markov chain sample of length $N_{it}\tau_h$ gives asymptotically the same Monte Carlo variance as an i.i.d. sample of size N_{it}; that is, we require τ_h dependent samples for each independent sample. Hence, the value of τ_h quantifies how well a Markov chain is exploring the state space, which, in alternative, can be qualitatively found through visual inspection. The smaller τ_h, the better the Markov chain is mixing. Large values of τ_h, on the other hand, correspond to the fact that the chain tends to be stuck at the same value for many iterations (because of consecutive rejections), or to typically make small moves.

In general, τ_h cannot be computed directly, and it is estimated by its sample version

$$\hat{\tau}_h := 1 + 2\sum_{k=1}^{N_{it}} \hat{\rho}_h(k)$$

where $\hat{\rho}_h(k)$, $|k| \in \{0, \dots, N_{it}\}$, is the sample autocorrelation function at lag k, ACF(k), defined as

$$\hat{\rho}_h(k) := \frac{\sum_{i=1}^{N_{it}-k}(h(X_i) - \bar{h})(h(X_{i+k}) - \bar{h})}{\sum_{i=1}^{N_{it}}(h(X_i) - \bar{h})^2}, \tag{3.15}$$

where $\bar{h} := \mathbb{E}_{\hat{p}}[h(X)]$ for simplicity of notation. The autocorrelation function and its sample version are symmetric, and take maximum value at the origin, while they are slowly but, usually, non-monotonically, decaying, as the lag increases. Thus, an indication of the chain mixing alternative to the IACT is how quickly the ACF decays with respect to the lag.

Finally, in order to obtain nearly independent samples, it is of common practice to sub-sample the chains by extracting every l^{th} element, such that $|\hat{\rho}(k)| < \epsilon$ for $k > l$, where ϵ is a fixed threshold.

3.2 Bayesian Parameter Inference for LV Models

We consider the case when the target of a MCMC sampler is the posterior distribution of the parameters $p(\boldsymbol{\theta}|\boldsymbol{x}) \propto p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})$ in the Bayesian setting, and the likelihood has integral representation (3.3), as in the LV models presented for the α-stable distribution. Observe that all the LVs models examined imply a marginal representation of the PDF but the structure of the integrand function of (2.24) is different from that of (2.27) and (2.39), making different inference methods suitable in each case. Observe furthermore that, in LV models, there is no conceptual difference between the parameters of a distribution and the latent variables involved in its representation, because they are both unknown quantities, but we are typically interested in inferring only the parameters.

3.2.1 Marginal Schemes

In the above scenario, producing *marginal* Markov chains, namely chains only in the original parameter space Θ, is challenging. In fact, the acceptance probability (3.12) of a MH kernel has expression

$$\begin{aligned}\rho &= 1 \wedge \frac{p(\boldsymbol{x}|\boldsymbol{\theta'})p(\boldsymbol{\theta'})}{p(\boldsymbol{x}|\boldsymbol{\theta}^{(k-1)})p(\boldsymbol{\theta}^{(k-1)})} \frac{q(\boldsymbol{\theta}^{(k-1)}|\boldsymbol{\theta'})}{q(\boldsymbol{\theta'}|\boldsymbol{\theta}^{(k-1)})} \\ &= 1 \wedge \frac{\int_{\mathbf{Y}} p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta'})\, \mathrm{d}\boldsymbol{y}\; p(\boldsymbol{\theta'})}{\int_{\mathbf{Y}} p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta}^{(k-1)})\, \mathrm{d}\boldsymbol{y}\; p(\boldsymbol{\theta}^{(k-1)})} \frac{q(\boldsymbol{\theta}^{(k-1)}|\boldsymbol{\theta'})}{q(\boldsymbol{\theta'}|\boldsymbol{\theta}^{(k-1)})},\end{aligned} \tag{3.16}$$

where $\boldsymbol{\theta}'$ is a sample from the proposal distribution $q(\boldsymbol{\theta}'|\boldsymbol{\theta}^{(k-1)})$, given the current state of the chain $\boldsymbol{\theta}^{(k-1)}$. It is often not possible to compute ρ analytically, because the integrals involved in (3.16) do not have a closed form expression. Accurate approximations of the acceptance ratio are possible sometimes, either through numerical approximation of the integrals involved, when this is not prohibitively costly, or by using approximate models that make the integrals tractable. However, this introduces bias in the sample posterior distribution wrt the true posterior, and in the Monte Carlo estimates computed. It is then not recommended the use marginal sampling schemes, unless the amount of error introduced by such approximation can be assessed and controlled.

3.2.2 Conditional Schemes

On the other hand, it is possible to produce exact samples from the posterior of interest, $p(\boldsymbol{\theta}|\boldsymbol{x})$, by targeting the augmented state space $\Theta' = \Theta \times \mathbf{Y}$ that incorporates the LVs $\boldsymbol{y}$. In fact, $p(\boldsymbol{\theta}|\boldsymbol{x})$ is the marginal of the posterior of the augmented state $p(\boldsymbol{\theta}, \boldsymbol{y}|\boldsymbol{x}) \propto p(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})p(\boldsymbol{\theta}, \boldsymbol{y}) = p(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$, meaning that the sample posterior of the parameters is reconstructed by simply discarding the LVs from the joint samples. The new augmented posterior can be evaluated because it involves the *integrand* of (3.3) and the acceptance probability is

$$\rho = 1 \wedge \frac{p(\boldsymbol{x}, \boldsymbol{y}'|\boldsymbol{\theta}')\, p(\boldsymbol{\theta}')}{p(\boldsymbol{x}, \boldsymbol{y}^{(k-1)}|\boldsymbol{\theta}^{(k-1)})\, p(\boldsymbol{\theta}^{(k-1)})} \frac{q(\boldsymbol{y}^{(k-1)}, \boldsymbol{\theta}^{(k-1)}|\boldsymbol{y}', \boldsymbol{\theta}')}{q(\boldsymbol{y}', \boldsymbol{\theta}'|\boldsymbol{y}^{(k-1)}, \boldsymbol{\theta}^{(k-1)})},$$

where, with obvious notation, we are now proposing, and either accepting or rejecting, jointly $(\boldsymbol{y}', \boldsymbol{\theta}')$ given $(\boldsymbol{y}^{(k-1)}, \boldsymbol{\theta}^{(k-1)})$. However, the new target state space is high dimensional and it will be typically difficult to design proposals $q(\boldsymbol{y}', \boldsymbol{\theta}'|\boldsymbol{y}^{(k-1)}, \boldsymbol{\theta}^{(k-1)})$ that efficiently sample the LVs and the parameters, even taking into account the factorization that derives from independence of the data and the respective LVs.

The standard solution is then to proceed via a GS *conditional* scheme that, at each iteration $k = 1, \ldots, N_{it}$ of the Markov chain, draws from the posterior full-conditionals

$$p(\boldsymbol{y}^{(k)}|\boldsymbol{\theta}^{(k-1)}, \boldsymbol{x}) = \prod_{n=1}^{N} p(y_n^{(k)}|\boldsymbol{\theta}^{(k-1)}, \boldsymbol{x}), \tag{3.17}$$

$$p(\theta_i^{(k)}|\boldsymbol{\theta}_{-i}^{(k)}, \boldsymbol{y}^{(k)}, \boldsymbol{x}), \qquad i = 1, \ldots, d, \tag{3.18}$$

where we use the notation $\boldsymbol{y}^{(k)} = [y_1^{(k)}, \ldots, y_N^{(k)}]$, $\boldsymbol{\theta}^{(k)} = [\theta_1^{(k)}, \ldots, \theta_d^{(k)}]$ and $\boldsymbol{\theta}_{-i}^{(k)} := [\theta_1^{(k)}, \ldots, \theta_{i-1}^{(k)}, \theta_{i+1}^{(k-1)}, \ldots, \theta_d^{(k-1)}]$, and we are now making explicit also the dependence on the data $\boldsymbol{x}$. Observe that the notation used here is in line with that of equation

(3.13) and in Algorithm 3: $\boldsymbol{y}^{(k)}$ is the N-dimensional vector of LVs at iteration k; $\boldsymbol{\theta}^{(k)}$ is the d-dimensional parameter vector at iteration k; $\boldsymbol{\theta}_{-i}^{(k)}$ is a $(d-1)$-dimensional parameter vector, with the first $i-1$ elements updated at iteration k, and the rest of the elements retained from the previous iteration, $k-1$; finally $y_n^{(k)}$ and $\theta_i^{(k)}$ are the n-th and i-th components of $\boldsymbol{y}^{(k)}$ $\boldsymbol{\theta}^{(k)}$, respectively. Observe, furthermore, that the factorization of (3.17) corresponds to conditional independence of the LVs given the data.

In this scenario it is either possible to sample from (3.17) and (3.18) or to target these distributions using a MH-GS scheme, in which the acceptance probabilities can again be evaluated, because the targets are

$$\begin{aligned} p(\boldsymbol{y}^{(k)}|\boldsymbol{\theta}^{(k-1)}, \boldsymbol{x}) &= \frac{p(\boldsymbol{x}, \boldsymbol{y}^{(k)}|\boldsymbol{\theta}^{(k-1)})}{p(\boldsymbol{x}|\boldsymbol{\theta}^{(k-1)})} \\ &\propto p(\boldsymbol{x}, \boldsymbol{y}^{(k)}|\boldsymbol{\theta}^{(k-1)}), \end{aligned}$$

and, for $i = 1, \ldots, d$,

$$p(\theta_i^{(k)}|\boldsymbol{\theta}_{-i}^{(k)}, \boldsymbol{y}^{(k)}, \boldsymbol{x}) = \frac{p(\boldsymbol{x}, \boldsymbol{y}^{(k)}|\theta_i^{(k)}, \boldsymbol{\theta}_{-i}^{(k)}) p(\theta_i^{(k)}|\boldsymbol{\theta}_{-i}^{(k)})}{p(\boldsymbol{x}, \boldsymbol{y}^{(k)}|\boldsymbol{\theta}_{-i}^{(k)})} \tag{3.19}$$

$$\propto p(\boldsymbol{x}, \boldsymbol{y}^{(k)}|\theta_i^{(k)}, \boldsymbol{\theta}_{-i}^{(k)}) p(\theta_i^{(k)}, \theta_{-i}^{(k)}), \tag{3.20}$$

and both expressions only involve the integrand of (3.3). It is typical, however, for conditional samplers to have a slow mixing of the chains: due to high posterior correlation existing between $\boldsymbol{\theta}$ and $\boldsymbol{y}$, the chains are likely to present no or very small movements for most of the iterations, with seldom large jumps.

A first way to improve the mixing is to aim at finding a transformation of the LVs or the parameters that reduces such correlation. The distribution of the transformed variables is found according to calculus rules (involving the Jacobian of the transformation), while the Markov chains on the original variables are obtained by simply inverse transforming the chains on the new variables. However, despite the existence of guidelines, the choice of the transformation to use depends on the LV model at hand. Moreover, if highly non-linear transformations and numerical inversion methods are involved in the process, this introduces approximations that destroy the purpose of exact sampling of this data augmented method.

3.2.3 Pseudo-Marginal Schemes

Recalling that in the marginal methods the LVs are ideally integrated out, overcoming the slow chain mixing problem and 'ad hoc' transformations, a solution to the posterior

parameter sampling is to mimic a marginal sampler, but treating the likelihood that appears in the acceptance probability (3.16) as a RV. When the likelihood is opportunely estimated, an unbiased estimate of the acceptance probability is produced. The methods based on this principle are called *pseudo-marginal* (PM) samplers or *exact approximations*,[13] and they combine the points of strength of the marginal and conditional schemes.

We start observing that we can re-write the likelihood (3.4) as

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{n=1}^{N} \int_{\mathcal{Y}_n} p(x_n, y_n|\boldsymbol{\theta})\, \mathrm{d}y_n = \prod_{n=1}^{N} \int_{\mathcal{Y}_n} \frac{p(x_n|y_n, \boldsymbol{\theta})p(y_n|\boldsymbol{\theta})}{q(y_n|\boldsymbol{\theta}, x_n)} q(y_n|\boldsymbol{\theta}, x_n)\, \mathrm{d}y_n,$$

where $q(y_n|\boldsymbol{\theta}, x_n)$ is an importance distribution for the LVs, that, for simplicity, we assume has support $\mathcal{Y}_n$. Then a Monte Carlo importance sampling based approximation of the likelihood is

$$p(\boldsymbol{x}|\boldsymbol{\theta}) \approx \hat{p}(\boldsymbol{x}|\boldsymbol{\theta}) := \prod_{n=1}^{N} \underbrace{\frac{1}{M} \sum_{j=1}^{M} \frac{p(x_n|Y_j, \boldsymbol{\theta})p(Y_j|\boldsymbol{\theta})}{q(Y_j|\boldsymbol{\theta}, x_n)}}_{\hat{p}(x_n|\boldsymbol{\theta})} = \prod_{n=1}^{N} \hat{p}(x_n|\boldsymbol{\theta}), \tag{3.21}$$

where, for each x_n, $n = 1, \dots N$, M samples $Y_j \sim q(y_n|\boldsymbol{\theta}, x_n)$, $j = 1, \dots, M$ are drawn from the importance distribution. With obvious notation, the M summands per stable data-point x_n, in (3.21),

$$W_j := \frac{p(x_n|Y_j, \boldsymbol{\theta})p(Y_j|\boldsymbol{\theta})}{q(Y_j|\boldsymbol{\theta}, x_n)} = \frac{p(Y_j|x_n, \boldsymbol{\theta})}{q(Y_j|\boldsymbol{\theta}, x_n)} p(x_n|\boldsymbol{\theta}), \qquad j = 1, \dots, M \tag{3.22}$$

are called *weights*.[14] The procedure is summarized in Algorithm 5. Notice that, by construction, (3.21) is a nonnegative unbiased estimator for $p(\boldsymbol{x}|\boldsymbol{\theta})$, namely $\hat{p}(\boldsymbol{x}|\boldsymbol{\theta}) \geq 0$ almost surely, and

$$\mathbb{E}\left[\hat{p}(\boldsymbol{x}|\boldsymbol{\theta})\right] = \prod_{n=1}^{N} \mathbb{E}_{q(y_n|\boldsymbol{\theta}, x_n)} \left\{ \frac{1}{M} \sum_{j=1}^{M} \frac{p(x_n|Y_j, \boldsymbol{\theta})p(Y_j|\boldsymbol{\theta})}{q(Y_j|\boldsymbol{\theta}, x_n)} \right\} \tag{3.23a}$$

$$= \prod_{n=1}^{N} \frac{1}{M} \sum_{j=1}^{M} \int_{\mathcal{Y}_n} p(x_n, y_n|\boldsymbol{\theta})\, \mathrm{d}y_n \tag{3.23b}$$

$$= \prod_{n=1}^{N} p(x_n|\boldsymbol{\theta}) = p(\boldsymbol{x}|\boldsymbol{\theta}). \tag{3.23c}$$

[13] This terminology refers to the fact that the posterior distribution of the parameters is exactly the stationary distribution of the Markov chain, despite the method makes use of an approximation of the likelihood in evaluating the acceptance probability.

[14] A more appropriate notation for the weights would be $\{W(Y_j|x_n, \boldsymbol{\theta})\}_{j=1}^{M}$, but we prefer $\{W_j\}_{j=1}^{M}$ for simplicity.

Algorithm 5 IS($\boldsymbol{\theta}, \boldsymbol{x}$) - IS based Monte Carlo estimate of the likelihood

1: Given $\boldsymbol{x} = [x_1, \dots, x_N]$ and $\boldsymbol{\theta}$
2: **for** $n = 1, \dots, N$ **do**
3: simulate $Y_j \sim q(y_n|\boldsymbol{\theta}, x_n)$, for $j = 1, \dots, M$
4: compute $W_j = \dfrac{p(x_n|Y_j, \boldsymbol{\theta})p(Y_j|\boldsymbol{\theta})}{q(Y_j|\boldsymbol{\theta}, x_n)}$, for $j = 1, \dots, M$
5: compute $\hat{p}(x_n|\boldsymbol{\theta}) = \frac{1}{M}\sum_{j=1}^{M} W_j$
6: **return** $\hat{p}(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{n=1}^{N} \hat{p}(x_n|\boldsymbol{\theta})$

The first attempt at using this idea was proposed in the 'Monte Carlo within Metropolis scheme' by ONeill et al. [2000], where a Monte Carlo estimate of the likelihood was independently produced twice for the acceptance probability (3.16) at each MCMC step, once for $\boldsymbol{\theta} = \boldsymbol{\theta}'$ at the numerator and once for $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k-1)}$ at the denominator. The estimate of the acceptance rate becomes then

$$\hat{\rho} := 1 \wedge \frac{\hat{p}(\boldsymbol{x}|\boldsymbol{\theta}')p(\boldsymbol{\theta}')}{\hat{p}(\boldsymbol{x}|\boldsymbol{\theta}^{(k-1)})p(\boldsymbol{\theta}^{(k-1)})} \frac{q(\boldsymbol{\theta}^{(k-1)}|\boldsymbol{\theta}')}{q(\boldsymbol{\theta}'|\boldsymbol{\theta}^{(k-1)})}. \tag{3.24}$$

However, the ratio of unbiased quantities is not unbiased for the ratio, and hence this method is approximate (in fact, a debiasing correction of $\hat{\rho}$ is proposed in the same paper).

Subsequent work by Beaumont [2003], and its further generalization by Andrieu and Roberts [2009],[15] showed that, by considering the augmented state $(\boldsymbol{\theta}, \zeta)$, where ζ is a nonnegative, unbiased estimate of the likelihood, it is possible to construct a sampler whose marginal distribution is *exactly* the posterior distribution of the parameters. In particular, let $p(\boldsymbol{\theta}, \zeta) = p(\zeta|\boldsymbol{\theta})p(\boldsymbol{\theta})$ be the distribution of the augmented state. It can be shown that an exact MCMC scheme for the target distribution proportional to $\zeta p(\zeta|\boldsymbol{\theta})p(\boldsymbol{\theta})$ has acceptance probability

$$\begin{aligned}\rho &= 1 \wedge \frac{\zeta' p(\zeta'|\boldsymbol{\theta}')p(\boldsymbol{\theta}')}{\zeta^{(k-1)} p(\zeta^{(k-1)}|\boldsymbol{\theta}^{(k-1)})p(\boldsymbol{\theta}^{(k-1)})} \frac{p(\zeta^{(k-1)}|\boldsymbol{\theta}^{(k-1)})q(\boldsymbol{\theta}^{(k-1)}|\boldsymbol{\theta}')}{p(\zeta'|\boldsymbol{\theta}')q(\boldsymbol{\theta}'|\boldsymbol{\theta}^{(k-1)})} \\ &= 1 \wedge \frac{\zeta' p(\boldsymbol{\theta}')}{\zeta^{(k-1)} p(\boldsymbol{\theta}^{(k-1)})} \frac{q(\boldsymbol{\theta}^{(k-1)}|\boldsymbol{\theta}')}{q(\boldsymbol{\theta}'|\boldsymbol{\theta}^{(k-1)})}.\end{aligned} \tag{3.25}$$

Thanks to the unbiasedness of ζ, marginalizing the samples to the $\boldsymbol{\theta}$ component corresponds to drawing from the desired target parameter posterior $p(\boldsymbol{\theta}|\boldsymbol{x})$, because

$$\int_{\mathbb{R}^+} \zeta p(\zeta|\boldsymbol{\theta})p(\boldsymbol{\theta})\, \mathrm{d}\zeta = \mathbb{E}_{p(\zeta|\boldsymbol{\theta})}[\zeta]p(\boldsymbol{\theta}) = p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}).$$

[15] The terminology *pseudo-marginal* sampler is used here for the first time.

Algorithm 6 IS-based pseudo-marginal MH

1: Given $\boldsymbol{x}$
2: set $\boldsymbol{\theta}^{(0)}$ arbitrarily
3: compute $\zeta^{(0)} \leftarrow \mathrm{IS}(\boldsymbol{\theta^{(0)}}, \boldsymbol{x})$
4: **for** $k = 1, \dots, N_{it}$ **do**
5: sample $\boldsymbol{\theta'} \sim q(\cdot|\boldsymbol{\theta}^{(k-1)})$
6: compute $\zeta' \leftarrow \mathrm{IS}(\boldsymbol{\theta'}, \boldsymbol{x})$
7: compute $\rho = 1 \wedge \dfrac{\zeta'}{\zeta^{(k-1)}} \dfrac{p(\boldsymbol{\theta'})}{p(\boldsymbol{\theta}^{(k-1)})} \dfrac{q(\boldsymbol{\theta}^{(k-1)}|\boldsymbol{\theta'})}{q(\boldsymbol{\theta'}|\boldsymbol{\theta}^{(k-1)})}$
8: set $\left\{\boldsymbol{\theta}^{(k)}, \zeta^{(k)}\right\} := \begin{cases} \{\boldsymbol{\theta'}, \zeta'\} & \text{with probability } \rho \\ \left\{\boldsymbol{\theta}^{(k-1)}, \zeta^{(k-1)}\right\} & \text{otherwise} \end{cases}$

Both conditions required for ζ are satisfied by the the Monte Carlo average $\hat{p}(\boldsymbol{x}|\boldsymbol{\theta})$ (3.21), that is thus a valid likelihood estimator in the PM sampler, as summarized in Algorithm 6.[16] In this case, the acceptance probability (3.25) has formal expression identical to (3.24), but there is a fundamental difference: in (3.25), at each step only one estimate of the likelihood (ζ') is required, $\zeta^{(k-1)}$ being retained from the previous iteration, while in (3.24) both quantities are estimated at each iteration.

Convergence of PM Schemes

Considering (3.25), we remark that, at each step k, we are either accepting or rejecting a move to the augmented proposed state $(\boldsymbol{\theta'}, \zeta')$ from the previous state $(\boldsymbol{\theta}^{(k-1)}, \zeta^{(k-1)})$ and that the probability of accepting this move is inversely proportional to $\zeta^{(k-1)}$, the noisy realization of the likelihood in the previous set of parameters. This is a critical point for the implementation of the pseudo-marginal method. Specifically, if at some iteration k we accept a proposed value of the parameters with a large positive noisy realization in the likelihood estimate

$$\zeta^{(k)} \gg p(\boldsymbol{x}|\boldsymbol{\theta}^{(k)}), \tag{3.26}$$

then the algorithm has the tendency of getting stuck at $\boldsymbol{\theta}^{(k)}$ for many iterations. Given that ζ is an unbiased estimator, (3.26) happens in practice when the variance of ζ is large, and ζ often takes values both smaller and larger than its expected value, $p(\boldsymbol{x}|\boldsymbol{\theta})$.

In particular, in our framework of Monte Carlo IS based estimate of the likelihood $\zeta = \hat{p}(\boldsymbol{x}|\boldsymbol{\theta})$ as in (3.21), the unbiasedness of ζ corresponds to the fact that, if we

[16] Observe that we are using a simplified notation wrt the literature, by considering only the RV corresponding to the likelihood estimator, instead of tracking all the LVs that are used to construct it via IS.

define

$$R(\boldsymbol{x}, \{Y_j\}_{j=1}^M) := \frac{\hat{p}(\boldsymbol{x}|\boldsymbol{\theta})}{p(\boldsymbol{x}|\boldsymbol{\theta})} = \prod_{n=1}^N \frac{\hat{p}(x_n|\boldsymbol{\theta})}{p(x_n|\boldsymbol{\theta})} = \prod_{n=1}^N \frac{\frac{1}{M}\sum_{j=1}^M W_j}{p(x_n|\boldsymbol{\theta})} \tag{3.27}$$

where W_j are given in (3.22), then $\mathbb{E}[R(\boldsymbol{x}, \{Y_j\}_{j=1}^M)] = 1$.[17] The estimator ζ has small variance when $R(\boldsymbol{x}, \{Y_j\}_{j=1}^M) \approx 1$ for any random realization of the weights W_j, while ζ has large variance when it frequently happens that $R(\boldsymbol{x}, \{Y_j\}_{j=1}^M) \approx 0$ or $R(\boldsymbol{x}, \{Y_j\}_{j=1}^M) \gg 1$. Given the definition (3.22) of the weights W_j, we have[18]

- $R(\boldsymbol{x}, \{Y_j\}_{j=1}^M) \approx 0$, if there is a data point x_n, $n = 1, \dots, N$, for which $\hat{p}(x_n|\boldsymbol{\theta}) \approx 0$. Equivalently, all the weights relative to the point x_n are approximately null, $W_j \approx 0$, $j = 1, \dots, M$, namely for all the samples Y_j proposed from $q(y_n|\boldsymbol{\theta}, x_n)$ we have $p(Y_j|x_n, \boldsymbol{\theta}) \approx 0$ and $q(Y_j|x_n, \boldsymbol{\theta}) \gg 0$;

- $R(\boldsymbol{x}, \{Y_j\}_{j=1}^M) \gg 1$, if there is at least one data point x_n, $n = 1, \dots, N$, for which $\hat{p}(x_n|\boldsymbol{\theta}) \gg 0$. Equivalently, for the data point x_n, there is at least one weight $W_j \gg 1$, namely one variable Y_j proposed from $q(y_n|\boldsymbol{\theta}, x_n)$, for which $p(Y_j|x_n, \boldsymbol{\theta}) \gg q(Y_j|x_n, \boldsymbol{\theta})$.

- $R(\boldsymbol{x}, \{Y_j\}_{j=1}^M) \approx 1$ if, for every data point x_n, each of the weights $W_j \approx p(x_n|\boldsymbol{\theta})$, $j = 1, \dots, M$. This is verified when the proposal distribution of the LVs $q(y_n|\boldsymbol{\theta}, x_n)$ is 'close' to their conditional distribution $p(y_n|\boldsymbol{\theta}, x_n)$.

Thus, to avoid poor mixing of the chains in the IS-based PM Algorithm 6, it is necessary to make the importance sampling produce likelihood estimates with small variance, and this is achieved by accurately choosing $q(y_n|\boldsymbol{\theta}, x_n)$, the importance distribution for the latent variables. The design of a proposal distribution that gives good mixing properties in the PM Algorithm 6 for the parameters of the α-stable distribution is our main contribution in Chapter 4.

3.2.4 Expectation Maximization for the MAP Estimator

The previous methods allow for the construction of the sample posterior distribution of the parameters. However, if in a LV model we are interested only in estimating the mode of the posterior distribution, (3.6), or maximum a posteriori (MAP), we

[17] Recall that the expectation is taken wrt the proposal distribution of the variables Y_j involved in the weights W_j.

[18] For simplicity, in the scenarios depicted below, we assume that all the factors of $\hat{p}(\boldsymbol{x}|\boldsymbol{\theta})$ but one have similar order of magnitude. This is a simplification, and numerical expedients need to be applied in practice, because the product of very large with very small factors $\hat{p}(x_n|\boldsymbol{\theta})$ frequently occurs; we refer to Chapter 4 for more details.

can make use the expectation-maximization algorithm (EM). This was introduced by Dempster et al. [1977] for maximum likelihood estimation in a LV (e.g. missing data) framework; see also Robert and Casella [2004]. However, EM can easily be adapted to MAP estimation, as proposed by Tanner [1997]. The method consists of two steps that are iterated until convergence. Given an initial guess of the parameters $\boldsymbol{\theta}^{(0)}$, at each iteration $k \geq 1$ we repeat:

(i) **E-step**: compute a lower bound of the log-posterior, given the data and the current estimate of the parameters, $\boldsymbol{\theta}^{(k-1)}$, [19]

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k-1)}) &:= \mathbb{E}_{p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}^{(k-1)})}[\log p(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta})] + \log p(\boldsymbol{\theta}) \\ &= \int_{\mathbf{Y}} \log p(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta}) p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}^{(k-1)})\, \mathrm{d}\boldsymbol{y} + \log p(\boldsymbol{\theta}) \\ &= \prod_{n=1}^{N} \int_{\mathcal{Y}_n} \log p(x_n, y_n|\boldsymbol{\theta}) p(y_n|x_n,\boldsymbol{\theta}^{(k-1)})\, \mathrm{d}y_n + \log p(\boldsymbol{\theta}), \end{aligned} \tag{3.28}$$

where $p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}^{(k-1)})$ is the conditional PDF of the LVs $\boldsymbol{y} = [y_1, \ldots, y_N]$, typically factorized by conditional independence, and $p(\boldsymbol{\theta})$ is the prior PDF of the parameters;

(ii) **M-step**: set the current value of the parameters to maximize $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k-1)})$

$$\boldsymbol{\theta}^{(k)} := \operatorname*{argmax}_{\boldsymbol{\theta}\in\Theta} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k-1)}). \tag{3.29}$$

When the integral in the E-step (3.28) cannot be computed exactly, a solution is to use a Monte-Carlo estimate for it. This leads to MCEM posterior parameter estimate for $\boldsymbol{\theta}$, introduced by Wei and Tanner [1990]. In particular, if we can draw $Y_1, \ldots, Y_M \sim p(y_n|x_n,\boldsymbol{\theta}^{(k-1)})$, $n = 1, \ldots, N$, then

$$Q_M(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k-1)}) := \prod_{n=1}^{N} \frac{1}{M} \sum_{j=1}^{M} \log p(x_n, Y_j|\boldsymbol{\theta}) + \log p(\boldsymbol{\theta}) \tag{3.30}$$

is an approximation of $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k-1)})$, that depends also on the sample size M used in the estimation of the integral. As in the EM algorithm, the M-step maximizes the lower bound on the posterior (3.30). We refer to Levine and Casella [2001] for implementations when it is not possible to draw samples from the conditional distribution of the LVs.

[19] The EM algorithm for maximum likelihood performs the same steps, but the log-prior does not appear in the lower bound $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k-1)})$.

3.3 Bayesian Parameter Inference for α-Stable Distributions

We now summarize how the above methods for parameter inference with LVs have been applied to the LV models (2.23), (2.27) and (2.39) for the α-stable distribution. The methods that we consider aim at the inference of the distribution parameters, and of the parameters of discrete-time linear models driven by stable noise. We mostly refer to scalar stable RVs, but the combination of the CMS with the SMiN model allows for inference for multivariate sub-Gaussian distributions.

3.3.1 Inference for the CMS LV Model

Based on representation (2.23), Buckle [1995] proposed a Gibbs sampler scheme for the parameters of generally skewed, scalar stable distributions $\boldsymbol{\theta} = (\alpha, \sigma, \beta, \mu)$. Given a stable dataset $\boldsymbol{x} = [x_1, \ldots, x_N]$, at the kth step the algorithm draws

$$y_n^{(k)} \sim p(y_n|\alpha^{(k-1)}, \beta^{(k-1)}, \sigma^{(k-1)}, \mu^{(k-1)}, x_n) \qquad n = 1, \ldots, N; \tag{3.31a}$$

$$\alpha^{(k)} \sim p(\alpha|\beta^{(k-1)}, \sigma^{(k-1)}, \mu^{(k-1)}, \boldsymbol{y}^{(k)}, \boldsymbol{x}); \tag{3.31b}$$

$$\beta^{(k)} \sim p(\beta|\alpha^{(k)}, \sigma^{(k-1)}, \mu^{(k-1)}, \boldsymbol{y}^{(k)}, \boldsymbol{x}); \tag{3.31c}$$

$$\sigma^{(k)} \sim p(\sigma|\alpha^{(k)}, \beta^{(k)}, \mu^{(k-1)}, \boldsymbol{y}^{(k)}, \boldsymbol{x}); \tag{3.31d}$$

$$\mu^{(k)} \sim p(\mu|\alpha^{(k)}, \beta^{(k)}, \sigma^{(k)}, \boldsymbol{y}^{(k)}, \boldsymbol{x}), \tag{3.31e}$$

where $\boldsymbol{y}$ is a vector of LVs, one per observation x_n. The LVs in (3.31a) are sampled through a RS algorithm based on adaptive proposals constructed in linear scale. The mechanism is similar to the ARS described in Section 3.1.1. However, rather than by log-concavity, the validity of the method is justified by the unimodality of $p(y_n|\alpha, \beta, \sigma, \mu, x)$ and the fact that its support $\mathcal{Y}_n$ is a compact set, making it possible to create a piece-wise uniform envelope proposal distribution.

A MH-GS is used for sampling α, β and μ in (3.31b), (3.31c), (3.31e), where a change of variables from $y_n^{(k)}$ to $v_n^{(k)} := t_{\alpha,\beta}(y_n^{(k)})$, or similar, is suggested, with $t_{\alpha,\beta}$ defined in (2.8).[20] This both adds computational complexity and it introduces bias in the sampling scheme. In fact, each time a new value of α, β or μ is proposed, it is necessary to evaluate the Jacobian of the inverse transformation $y = t_{\alpha,\beta}^{-1}(v)$: such inversion cannot be done analytically, making it necessary to use numerical schemes that introduce ideally small, but not quantifiable, errors. However, the new target full-conditionals $p(\alpha|\beta, \sigma, \mu, \boldsymbol{v}, \boldsymbol{x})$, $p(\beta|\alpha, \sigma, \mu, \boldsymbol{v}, \boldsymbol{x})$ and $p(\mu|\alpha, \beta, \sigma, \boldsymbol{v}, \boldsymbol{x})$ are more

[20] We omit here the subscript notation on the parameter β that refers to the parametrization of the stable distribution.

evenly spread on the support (less "peaky" and less oscillating) than the original ones, due to less correlation existing between the parameters and the latent variables, and the overall effect is a better mixing of the chains. The proposal distributions for the MH-GS scheme in Buckle [1995] are constructed using ARS (for non log-concave densities), even if we find that Gaussian random walk proposals with adequately tuned variance work well, see Chapter 4 for the details. Uniform priors are assumed: despite being an improper choice for σ and μ, this mitigates the need of modelling a priori the parameter dependence, and it still leads to a proper posterior distribution.

Finally, the parameter σ can be sampled exactly. In fact, an inverse gamma prior is conjugate to the likelihood for transformed parameter $\sigma^{\alpha/(\alpha-1)}$, and this leads to a posterior full conditional in the same family of distributions; σ is then obtained through a simple inverse transformation.

We provide simulation results of the above method in Chapter 4, and compare them with a PM sampler, based on Algorithm 6, that is the first contribution of this thesis. The two schemes achieve comparable chain mixing, but the PM sampler has the theoretical advantage of being asymptotically exact, without involving highly non-linear transformations, such as $t_{\alpha,\beta}$.

Based on the LV model (2.23), and on the parameter sampler proposed by Buckle [1995], Qiou and Ravishanker [1998a] provided a way for simultaneously estimating the distribution parameters $\boldsymbol{\theta} = (\alpha, \sigma, \beta, \mu)$, and the parameters of a linear autoregressive moving-average (ARMA) model driven by stable noise. This includes the AR(P) model introduced in Chapter 1.1, equations (1.1) and (1.2), that we consider here for simplicity, implying that the model parameters to be estimated coincide with $\boldsymbol{\lambda}$. Each noise term u_n, $n = 1, \ldots, N$ in (1.1) can be expressed as

$$u_n = x_n - \sum_{j=1}^{P} \lambda_j x_{n-j}, \tag{3.32}$$

where it is assumed that $u_n \overset{i.i.d.}{\sim} \mathcal{S}_\alpha(\sigma, \beta, 0)$, to model the extreme behaviour of sums of effect, according to the generalized CLT. This implies that $p(\mathbf{x}|\boldsymbol{\lambda}, \alpha, \sigma, \beta, \mu)$, the likelihood of the data $\mathbf{x} = \{x_n\}_{n=1}^N$, is the product of N stable densities, each with marginal expression of the form (2.23). Then a Gibbs sampler can be devised, that at each iteration draws $\boldsymbol{y}$, α, σ, $[\beta, \mu, \boldsymbol{\lambda}]$ from their respective posterior full conditionals. Here $\boldsymbol{y}$ is a vector of LVs, one per noise term, or, equivalently, one per observation x_n. The variables $\boldsymbol{y}$, α and σ are sampled in a similar way to Buckle [1995]. The autoregressive parameters $\boldsymbol{\lambda}$ are sampled in block with β and μ, via a MH within Gibbs step with Gaussian random walk proposal, where the proposed

parameters are accepted only if they make the AR(P) system stationary, namely if the roots of the polynomial

$$z^P - \sum_{j=1}^{P} \lambda_j z^{P-j}$$

all lie inside the unit circle: $|z_j| < 1$, $j = 1, \ldots, P$. Qiou and Ravishanker [1998a] discuss also the adequacy of the α-stable distribution for $\{u_n\}$ in ARMA models, as well as model selection (namely the choice of P, based on the observations).

Despite the successful results shown, this sampler involves again highly non-linear transformations and numerical inversions, unlike the schemes based on the SMiN and the approximate MaSMiN representations.

3.3.2 Inference for the SMiN LV Model

The conditionally Gaussian representation (2.26a)-(2.26b) of symmetric stable RVs allows to extend inference methods that are available for Gaussian distributions, by augmenting the state to include the mixing variable S.

Here we summarize, for example, the samplers proposed at the same time, but independently, by Godsill [1999], Godsill and Kuruoğlu [1999], and Tsionas [1999], for the joint estimation of the parameters $\boldsymbol{\lambda}$ of the discrete time linear model (1.1), when $u_n \overset{i.i.d.}{\sim} \mathcal{S}_\alpha(\sigma, 0, 0)$, and of the distribution parameters α and σ of the driving noise process. Let $\boldsymbol{s} := [s_1, \ldots, s_N]$ be the vector of latent variables, with each

$$s_n \overset{i.i.d.}{\sim} \mathcal{S}_{\alpha/2}((\cos(\pi\alpha/4))^{2/\alpha}, 1, 0) \tag{3.33}$$

corresponding to a noise term u_n, according to the SMiN representation for symmetric stable distributions. Then[21]

$$u_n|\alpha, \sigma, s_n \sim \mathcal{N}(0, 2\sigma^2 s_n), \tag{3.34}$$

and, from (1.1), it follows that $p(\mathbf{x}|\boldsymbol{\lambda}, \alpha, \sigma, \boldsymbol{s})$, the likelihood of the data $\mathbf{x}$ conditioned also on the LVs $\boldsymbol{s}$, is Gaussian

$$\mathbf{x}|\boldsymbol{\lambda}, \alpha, \sigma, \boldsymbol{s} \sim \mathcal{N}(\mathbf{G}\boldsymbol{\lambda}, \tilde{\boldsymbol{\Sigma}}), \tag{3.35}$$

[21] Given the Bayesian framework, we now explicitly condition also on the distribution parameters α and σ.

where $\tilde{\boldsymbol{\Sigma}}$ is an $N \times N$ diagonal matrix, with diagonal elements $2\sigma^2 s_n$. Then a Gibbs sampler can be implemented, that, at each iteration k, samples

$$s_n^{(k)} \sim p(s_n|\alpha^{(k-1)}, \sigma^{(k-1)}, \boldsymbol{\lambda}^{(k-1)}, x_n) \qquad n = 1, \dots, N; \tag{3.36a}$$

$$\sigma^{(k)} \sim p(\sigma|\boldsymbol{s}^{(k)}, \alpha^{(k-1)}, \boldsymbol{\lambda}^{(k-1)}, \boldsymbol{x}); \tag{3.36b}$$

$$\boldsymbol{\lambda}^{(k)} \sim p(\boldsymbol{\lambda}|\boldsymbol{s}^{(k)}, \alpha^{(k-1)}, \sigma^{(k)}, \boldsymbol{x}); \tag{3.36c}$$

$$\alpha^{(k)} \sim p(\alpha|\boldsymbol{s}^{(k)}, \sigma^{(k)}, \boldsymbol{\lambda}^{(k)}, \boldsymbol{x}). \tag{3.36d}$$

Based on Bayes' theorem and (3.35), the posterior full conditionals (3.36b) and (3.36c) have the same structure as those of a heteroscedastic Gaussian linear model, see for example [Gelman et al., 2014, Chapter 14.8]. Specifically, a Gaussian prior for $\boldsymbol{\lambda}$ and inverse gamma for σ are conjugate to the conditional likelihood (3.35), making it straightforward to sample from these distributions, after a simple update of the prior parameters.

The conditionally Gaussian structure (3.35) allows to construct an exact sampler also for s_n. In fact, via Bayes' theorem,

$$p(s_n|\alpha, \sigma, \boldsymbol{\lambda}, x_n) \propto \mathcal{N}(0, 2\sigma^2 s_n) p(s_n), \qquad n = 1, \dots, N, \tag{3.37}$$

where $p(s_n)$ is the α-stable prior distribution (3.33), that can be sampled exactly using the CMS method. This allows one to choose the prior $p(s_n)$ as the proposal distribution in a MH-GS step with independent proposals, leading to an acceptance probability of the new value equal to

$$\rho = 1 \wedge \frac{\mathcal{N}(0, 2\sigma^2 s_n')}{\mathcal{N}(0, 2\sigma^2 s_n)}, \tag{3.38}$$

which can be exactly evaluated. Alternatively, given that the Gaussian likelihood is maximized when $s_n = x_n^2/2\sigma^2$, yielding the bound

$$\mathcal{N}(0, 2\sigma^2 s_n) \leq \frac{1}{\sqrt{2\pi x_n^2}} \exp(-1/2), \tag{3.39}$$

a rejection sampler can be implemented, using again the stable prior as proposal for s_n. In fact, with this choice of the proposal, (3.39) translates into a sensible acceptance criterion for the RS scheme. As observed in Godsill [1999], however, the RS can present low acceptance rate, or, accordingly, the MH can move slowly around the state space, in correspondence of an either very large or very small observation x_n. For these cases, Godsill [1999] and Godsill and Kuruoğlu [1999] propose an approximate hybrid RS with improved acceptance rate, that either

approximates the prior $p(s_n)$ or the likelihood $\mathcal{N}(0, 2\sigma^2 s_n)$. Finally, Godsill [2000] proposes to use a slice sampling scheme, by further augmenting the state space, motivated by the fact that this could have better convergence properties than the MH or RS methods.

The step (3.36d) for sampling α is the most difficult one, because the target is not in standard form. Godsill [1999], Godsill and Kuruoğlu [1999], and Godsill [2000] consider α to be known, while Tsionas [1999] proposes an approximate sampler. Instead of targeting the conditional distribution $p(\alpha|\boldsymbol{s}^{(k)}, \sigma^{(k)}, \boldsymbol{\lambda}^{(k)}, \boldsymbol{x})$, Tsionas [1999] targets $p(\alpha|\sigma^{(k)}, \boldsymbol{\lambda}^{(k)}, \boldsymbol{x})$. According to (3.16), this is proportional to the intractable stable PDF, because the LVs are integrated out. The reason for this choice is that numerical methods (e.g. Fast Fourier Transforms) are used in both cases, but sampling from $p(\alpha|\sigma^{(k)}, \boldsymbol{\lambda}^{(k)}, \boldsymbol{x})$ appears to give smaller numerical error as well as smaller computation time. A MH kernel with Gaussian random walk proposal, and opportunely tuned variance is used for this approximate step.

In this same framework, Godsill [1999] and Godsill and Kuruoğlu [1999] propose also a MCEM algorithm to perform MAP inference on $\boldsymbol{\lambda}$ and σ (α is again assumed to be known). The expectation wrt the conditional distribution of the LVs $\boldsymbol{s}$ in the E-step is approximated through a Monte Carlo average, where the samples are obtained using either the MH or the (hybrid) RS, as above. The M-step has a form similar to that of Gaussian models, and the full algorithm can be interpreted as iteratively reweighted least squares scheme.

3.3.3 Inference Combining the SMiN and CMS LV Models

It is possible to combine the SMiN and CMS models and this idea was developed by Ravishanker and Qiou [1998], Qiou and Ravishanker [1998b], Ravishanker and Qiou [1999] for inference for sub-Gaussian random vectors $\mathbf{X}$, that have marginal representation of the PDF (2.28).[22] Given that S, the mixing RV in the SMiN representation, is positive stable, its PDF $p_S(\cdot)$ has marginal form (2.23), in which an auxiliary variable Y is associated to S through the CMS generative scheme. Then, combining the two models we have the following marginal represntation of a multivariate stable PDF

$$p_{\mathbf{X}}(\mathbf{x}) = \int_{\mathbb{R}^+} \int_{\mathcal{Y}} p_{(\mathbf{X},S,Y)}(\mathbf{x}, s, y)\, \mathrm{d}y\, \mathrm{d}s, \tag{3.40}$$

[22] This combined scheme can be used also for scalar RVs, but the simpler individual LV models are typically preferred in this case. Furthermore, there is no explicit indication in the literature to why this combined model is used for the multivariate stable distribution, instead of the SMiN representation (2.28), but our guess is that the reason is to improve efficiency of the sampler.

with $\mathcal{Y}$ as in (2.22), and where the PDF $p_{(\mathbf{X},S,Y)}(\mathbf{x}, s, y)$ of the augmented vector $(\mathbf{X}, S, Y)$ has closed form expression. Recall that, in this representation, the distribution parameters are $\alpha, \boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

As the target of inference, Ravishanker and Qiou [1998], Qiou and Ravishanker [1998b] consider VARMA (vector ARMA) models driven by a d-dimensional noise process $\{\mathbf{u}_n\}$ with symmetric sub-Gaussian stable distribution. VARMA models are used to characterize the dependence structure within and between multiple series. We focus on the VAR(P) subclass, for simplicity. In analogy with (3.32), for each of the d-valued observations $\mathbf{x}_n$, $n = 1, \ldots, N$, we can write

$$\mathbf{u}_n = \mathbf{x}_n - \sum_{j=1}^{P} \boldsymbol{\Lambda}_j \mathbf{x}_{n-j}, \tag{3.41}$$

where now each $\boldsymbol{\Lambda}_j$ is a $d \times d$-dimensional matrix. Then $p(\mathbf{x}|\{\boldsymbol{\Lambda}_j\}_{j=1}^P, \alpha, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, the likelihood of the data $\mathbf{x} = \{\mathbf{x}_n\}_{n=1}^N$, is the product of N terms that have marginal form (3.40). This enables sampling from the posterior of the distribution parameters $\alpha, \boldsymbol{\mu}, \boldsymbol{\Sigma}$ and of the VARMA model parameters $\{\boldsymbol{\Lambda}_j\}_{j=1}^P$, through a Gibbs sampler that augments the state space by incorporating the LVs $\boldsymbol{y} = [y_1, \ldots, y_N]$ and $\boldsymbol{s} = [s_1, \ldots, s_N]$. With some modifications caused by the combination of the SMiN with the CMS LV model,[23] the sampling steps for α and $\boldsymbol{y}$ are similar to those in Section 3.3.1. The variables $\boldsymbol{s}$ are sampled through the ratio of uniforms method introduced by Kinderman and Monahan [1977], Wakefield et al. [1991], and consisting of generating two uniform RVs on an appropriate domain and taking their ratio as the sample for each s_n; in this case, generation on the exact domain of interest in not possible, and Qiou and Ravishanker [1998b] propose an approximation based on a numerical search algorithm. Finally, the posterior conditional distribution of $\boldsymbol{\Sigma}^{-1}$ is a Wishart distribution, that of $\boldsymbol{\mu}$ is a multivariate Gaussian, and that of each element of each matrix $\boldsymbol{\Lambda}_j$ is Gaussian (with stationarity imposed through rejection), making these Gibbs steps straightforward.

Based on (3.40), Ravishanker and Qiou [1999] provide a MCEM algorithm for MAP point estimation of the distribution parameters $\alpha, \boldsymbol{\mu}, \boldsymbol{\Sigma}$ of sub-Gaussian vectors. A Gibbs sampler is used to draw the LVs $\boldsymbol{y}$ and $\boldsymbol{s}$ needed for the Monte Carlo estimate (3.30). In particular, at each EM step k, samples are drown from $p(\boldsymbol{y}|\boldsymbol{s}, \alpha^{(k-1)}, \boldsymbol{\mu}^{(k-1)}, \boldsymbol{\Sigma}^{(k-1)}, \boldsymbol{x})$ and $p(\boldsymbol{s}|\boldsymbol{y}, \alpha^{(k-1)}, \boldsymbol{\mu}^{(k-1)}, \boldsymbol{\Sigma}^{(k-1)}, \boldsymbol{x})$, through the same

[23] In fact, the scale parameter σ of the stable LV S is a deterministic function of α, and the full conditional distribution of α depends on the vector $\boldsymbol{s}$, that is updated at every GS step, making this distribution difficult to target. A MH with suitable proposals (linearly transformed Beta distributions) is used for this purpose.

mechanism used in Qiou and Ravishanker [1998b], and with the current estimate used for the distribution parameters.

3.3.4 Inference for the Approximate MaSMiN LV Model

As anticipated in Section 2.4.4, it is not possible to make use of the exact MaSMiN model. Both the GAA and the GAMA schemes have been used for exact inference for the stemming approximated stable distribution, namely for approximate inference for the exact stable distribution distribution. Recall that, when $W_j \sim \mathcal{N}(\mu_W, \sigma_W^2)$, the distribution parameters σ and β are re-parametrized to μ_W and σ_W, according to equations (2.34a) and (2.34b).

The GAMA scheme is used in Lemke [2014], Lemke and Godsill [2014] and Lemke et al. [2015] for inference of the stable distribution parameters. Using the GAMA approximation, the augmented state associated to each stable observation x_n, $n = 1, \dots, N$, is formed by:

- $\mathbf{\Gamma}^n_{(0,c)} := \left[\Gamma^{1,n}_{(0,c)}, \dots, \Gamma^{N^n_{(0,c)},n}_{(0,c)}\right]$, where $N^n_{(0,c)}$ the number of LVs $\{\Gamma_j\}$ generated, typically different (because random) for each x_n; these variables are used to approximate the first part of the series (2.36)-(2.37);
- $\mathbf{R}^n_{(c,\infty)} := (R^{1,n}_{(c,\infty)}, R^{2,n}_{(c,\infty)})$, a bivariate vector, whose role is to provide a summary of the tails of the series (2.36)-(2.37).

These variables are used to form the approximate mean $\hat{M}_2^{(n)}$ and variance $\widehat{S^2}_2^{(n)}$ for each observation, and, as a consequence, it is possible to re-write (2.45a) for each x_n as[24]

$$x_n|\alpha, \mu_W, \sigma_W, \mu, \mathbf{\Gamma}^n_{(0,c)}, \mathbf{R}^n_{(c,\infty)}, \overset{\text{approx}}{\sim} \mathcal{N}(\hat{m}_2^{(n)}, \widehat{s^2}_2^{(n)}), \tag{3.42}$$

where $\hat{m}_2^{(n)}$ and $\widehat{s^2}_2^{(n)}$ are the realizations of $\hat{M}_2$ and $\hat{S}_2^2$ for each observation. As mentioned in section 2.4.4, one of the advantages of the GAMA over the GAA scheme is that (μ_W, σ_W, μ) can be marginalized in (3.42), leading to a closed form expression for[25] the conditional distribution

$$x_n|\alpha, \mathbf{\Gamma}^n_{(0,c)}, \mathbf{R}^n_{(c,\infty)}, \tag{3.43}$$

as reported in the literature. A Gibbs sampler can then be devised, that makes use of either or (3.42) or of (3.43) to sample from the full conditionals of the parameters

[24] Given the Bayesian inference framework, we explicitly condition also on the other distribution parameters.

[25] This is also an approximation, given that it derives from (3.42).

and the LVs. In particular, given $\boldsymbol{x} = [x_1, \dots, x_N]$, and defining

$$\mathbb{T}_{(0,c)} := \{\boldsymbol{\Gamma}^n_{(0,c)}\}_{n=1}^N, \qquad \mathbb{R}_{(c,\infty)} := \{\mathbf{R}^n_{(c,\infty)}\}_{n=1}^N, \tag{3.44}$$

then, at each iteration k, the GS draws

$$(\mu_W^{(k)}, \mu^{(k)}, {\sigma_W^2}^{(k)}) \sim p(\mu_W, \mu, \sigma_W^2 | \alpha^{(k-1)}, \mathbb{T}^{(k-1)}_{(0,c)}, \mathbb{R}^{(k-1)}_{(c,\infty)}, \boldsymbol{x}); \tag{3.45a}$$

$$\alpha^{(k)} \sim p(\alpha | \mathbb{T}^{(k-1)}_{(0,c)}, \mathbb{R}^{(k-1)}_{(c,\infty)}, \boldsymbol{x}); \tag{3.45b}$$

$$(\boldsymbol{\Gamma}^n_{(0,c)})^{(k)} \sim p(\boldsymbol{\Gamma}^n_{(0,c)} | \mu_W^{(k)}, \mu^{(k)}, {\sigma_W^2}^{(k)}, \alpha^{(k)}, (\mathbf{R}^n_{(c,\infty)})^{(k-1)}, \boldsymbol{x}), \quad 1 \le n \le N; \tag{3.45c}$$

$$(\mathbf{R}^n_{(c,\infty)})^{(k)} \sim p(\mathbf{R}^n_{(c,\infty)} | \mu_W^{(k)}, \mu^{(k)}, {\sigma_W^2}^{(k)}, \alpha^{(k)}, (\boldsymbol{\Gamma}^n_{(0,c)})^{(k)}, \boldsymbol{x}), \quad 1 \le n \le N. \tag{3.45d}$$

The first step can be sampled exactly, observing that the conditional distribution in (3.45a) can be factorized as

$$p(\mu_W, \mu | \alpha^{(k-1)}, \sigma_W^{(k-1)}, \mathbb{T}^{(k-1)}_{(0,c)}, \mathbb{R}^{(k-1)}_{(c,\infty)}, \boldsymbol{x}) \times p(\sigma_W^2 | \alpha^{(k-1)}, \mathbb{T}^{(k-1)}_{(0,c)}, \mathbb{R}^{(k-1)}_{(c,\infty)}, \boldsymbol{x})$$

where, if uniform priors (improper for μ_W and μ) are assumed, the first factor is a Gaussian distribution for the vector (μ_W, μ), while the second is an inverse gamma distribution for σ_W^2; the form of both of these distributions is a direct consequence of the approximate conditionally Gaussian likelihood (3.42).

In the second step, the marginalized likelihood (3.43) is used; the conditional distribution in (3.45b) results to be proportional to (3.43), if a uniform prior for α is assumed. Then a MH-GS step can be performed to target the posterior conditional of α (Gaussian proposals are used in the literature). It has to be noted that the variables $\mathbf{R}^n_{(c,\infty)}$ are a function of α, given their role of summary statistics of the tails of the series (2.36)-(2.37), in which the exponent for each Γ_j is $-1/\alpha$. Thus, in order to evaluate the MH acceptance probability, when a new value α' is proposed, a new value $(\mathbf{R}^n_{(c,\infty)})'$ needs to be sampled too; we refer to Lemke et al. [2015] for the details.

For both steps (3.45c) and (3.45d), a MH-GS can be implemented, by choosing the proposals to coincide with the priors. The finite number of variables $\{\Gamma_j\}$ in each $\boldsymbol{\Gamma}^n_{(0,c)}$ can be exactly simulated as sum of exponentials, as explained in Section 2.4.3. On the other hand, the prior for $\mathbf{R}^n_{(c,\infty)}$ is approximated with a bivariate Gaussian distribution, according to the GAMA scheme; we refer to Lemke [2014] and Lemke et al. [2015] for the details. Then the target distributions in the MH steps are proportional to the likelihood (3.42) (or (3.43), if the marginalized version is used. Alternatively, and in analogy with the sampling step for the variables $\boldsymbol{s}$ in the SMiN scheme in equation (3.37), a RS can be performed for both $\boldsymbol{\Gamma}^n_{(0,c)}$ and $\mathbf{R}^n_{(c,\infty)}$. Such

RS is based on the fact that the approximate conditionally Gaussian likelihood (3.42) is bounded, similar to (3.39) in the SMiN, and this bound is useful when the RS draws are proposed from the priors. Finally, improvements of the acceptance ratios in correspondence of large observations x_n can be achieved by making use of the observations in Appendix A.3 of Chapter 2, and we refer directly to the literature for these.

The GAA method can be adopted whenever the marginalized likelihood (3.43) is not needed. For example, Lemke and Godsill [2012] use it for sampling from the posterior distribution of the parameters $\boldsymbol{\lambda}$ in an AR(P) model (1.1)-(1.2), driven by skewed stable noise u_n. The augmented state associated to each noise term u_n is formed by the vector $\boldsymbol{\Gamma}^n_{(0,c)}$, defined as in the section above, and $R^n_{(c,\infty)}$, a scalar, whose role is to summarize the PSR of each u_n, directly. The GAA scheme approximates $R^n_{(c,\infty)}$ with a Gaussian RV. Recall that, unlike the GAMA method, the RV $R^n_{(c,\infty)}$ does not need to be sampled, because only its moments are used in the definition of $\hat{M}_1^{(n)}$ and $\widehat{S^2}_1^{(n)}$, the moments of the approximate conditionally Gaussian representation of each noise term u_n[26]

$$u_n|\boldsymbol{\Gamma}^n_{(0,c)} \overset{\text{approx}}{\sim} \mathcal{N}(\hat{m}_1^{(n)}, \widehat{s^2}_1^{(n)}).$$

From this, it follows that the likelihood of the data $\boldsymbol{x}$, conditioned on the LVs $\mathbb{T}_{(0,c)}$, defined in (3.44), is also approximately Gaussian

$$\mathbf{x}|\boldsymbol{\lambda}, \mathbb{T}_{(0,c)} \overset{\text{approx}}{\sim} \mathcal{N}(\mathbf{G}\boldsymbol{\lambda} + \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}),$$

where $\tilde{\boldsymbol{\mu}}$ is an N-dimensional vector with elements $\hat{m}_1^{(n)}$, and $\tilde{\boldsymbol{\Sigma}}$ is an $N \times N$ diagonal matrix, with diagonal elements $\widehat{s^2}_1^{(n)}$. In analogy to what we presented for the SMiN model in Section 3.3.2, regular inference can then be carried out as for the Gaussian model, by augmenting the set of parameters to be estimated to $\{\boldsymbol{\lambda}, \mathbb{T}_{(0,c)}\}$. Assuming that the order of the model P is known, a Gibbs sampler can be devised, that at the k-th iteration draws

$$\boldsymbol{\lambda}^{(k)} \sim p(\boldsymbol{\lambda}|\mathbb{T}^{(k-1)}_{(0,c)}, \boldsymbol{x}), \tag{3.46a}$$

$$(\boldsymbol{\Gamma}^n_{(0,c)})^{(k)} \sim p(\boldsymbol{\Gamma}^n_{(0,c)}|\boldsymbol{\lambda}^{(k)}\boldsymbol{x}), \qquad n = 1, \ldots, N; \tag{3.46b}$$

In order to sample $\boldsymbol{\lambda}$ in (3.46a), a conjugate Gaussian prior can be adopted, leading to a Gaussian conditional posterior. The sampling technique for the variables $\boldsymbol{\Gamma}^n_{(0,c)}$

[26] Here the distribution parameters $\alpha, \mu_W, \sigma_W, \mu$ are assumed to be known.

in (3.46b) is conceptually the same as the one used for sampling from (3.45c) and either a MH or a RS step can be adopted.

Finally, we refer to [Lemke, 2014, Chapter 6] for simultaneous inference of the parameters and the order P of the autoregressive model, based on a reversible jump MCMC sampler, that exploits once more the approximate conditionally Gaussian structure of the likelihood stemming from the GAA scheme.

3.3.5 Summary

Table 3.1 summarizes the Bayesian inference methods for LV models for the stable distribution, that were presented in this Section. In particular, we outline the LV involved in each of the models used, and the parameter inference procedures that have been implemented in the literature. In this table we indicate with $\boldsymbol{\theta}$ the distribution parameters, corresponding to $\boldsymbol{\theta} = \{\alpha, \sigma, \beta, \mu\}$ in the CMS model, $\boldsymbol{\theta} = \{\alpha, \sigma\}$ in the SMiN, $\boldsymbol{\theta} = \{\alpha, \boldsymbol{\mu}, \boldsymbol{\Sigma}\}$ in the SMiN combined with the CMS model,[27] and $\boldsymbol{\theta} = \{\alpha, \mu_W, \sigma_W, \mu\}$ in the (approximate) MaSMiN model.

We want to underscore that this is not a comprehensive summary of all the available Bayesian inference methods existing in the literature. As anticipated in Section 1.3, we have not included Bayesian samplers that have a more approximate nature, such as that of Peters et al. [2011] and Peters et al. [2012], based on summary statistics, or that of Lombardi [2007], based on inversion of the characteristic function, given that our intention is to work on exact sampling methods. With respect to this, in Chapter 4 we provide an exact Bayesian sampler for the CMS LV model, based on pseudo-marginal schemes as in Section 3.2.3.

Observe that the inference procedures that we summarized for the MaSMiN model are also approximate, because they are exact for an approximate model. However, our contributions in Chapter 5, 6 and 7, is to quantify the amount of approximation introduced when using the MaSMiN model based on the GAA scheme, allowing for its systematic control through the selection of the truncation parameter c. This justifies our inclusion of the inference for the approximate MaSMiN models in our considerations.

We remark also that, in this thesis we have not tackled state inference methods in stochastic dynamics with stable noise process. We refer to Lombardi and Godsill [2006] and Lombardi and Godsill [2004] for state inference with the SMiN representation, and Lemke [2014] and Lemke and Godsill [2015] for inference with the MaSMiN model and the GAMA scheme. In Chapter 8 we rather work on simulation from multivariate stable distributions that are at the solution of linear SDEs driven

[27] Recall that this is used for multivariate sub-Gaussian distributions.

by stable Lévy processes, using the MaSMiN representation; this is the first step necessary for implementing Bayesian state inference.

Our work in Chapter 8 has also potential for parameter inference for multivariate distributions. In fact, it poses the basis for approximating the multivariate MaSMiN model to any multivariate stable distributions, regardless of the nature of its spectral measure function. This will enable to relax the assumptions of discrete spectral measure, or of sub-Gaussianity of the stable distribution, that are usually made in the literature.

Table 3.1 Summary of the LVs models and Bayesian posterior inference methods considered for $\boldsymbol{\theta}$, the α-stable distribution parameters (see the text); $\boldsymbol{\lambda}$, the parameters of a linear (possibly autoregressive) model (1.1)-(1.2) with stable error terms; $\boldsymbol{\Lambda}_j$, $j = 1, \ldots, P$, the matrices that parametrize the VAR(P) model (3.41) with multivariate stable error terms.

Model	LVs	Parameters	Methods
CMS	$[y_1, \ldots, y_N]$	$\boldsymbol{\theta}$	GS (ARS, MH-GS steps)
		$\{\boldsymbol{\theta}, \boldsymbol{\lambda}\}$	GS (ARS, MH-GS steps)
	$\{Y_{1j}\}_{j=1}^{M}, \ldots, \{Y_{Nj}\}_{j=1}^{M}$	$\boldsymbol{\theta}$	PM (ARS, MH-GS steps)
SMiN	$[s_1, \ldots, s_N]$	$\{\boldsymbol{\theta}, \boldsymbol{\lambda}\}$	GS (direct, RS or MH-GS steps)
SMiN + CMS	$[s_1, \ldots, s_N]$ $[y_1, \ldots, y_N]$	$\{\boldsymbol{\theta}, \boldsymbol{\Lambda}_j\}$	GS (MH-GS, ratio of uniforms, ARS, exact steps)
MaSMiN (GAMA)	$[\boldsymbol{\Gamma}^1_{(0,c)}, \ldots, \boldsymbol{\Gamma}^N_{(0,c)}]$ $[\mathbf{R}^1_{(c,\infty)}, \ldots, \mathbf{R}^N_{(c,\infty)}]$	$\boldsymbol{\theta}$	GS (direct, RS or MH-GS steps)
MaSMiN (GAA)	$[\boldsymbol{\Gamma}^1_{(0,c)}, \ldots, \boldsymbol{\Gamma}^N_{(0,c)}]$ $[R^1_{(c,\infty)}, \ldots, R^N_{(c,\infty)}]$	$\boldsymbol{\lambda}$	GS (direct, RS or MH-GS steps)

Chapter 4

Pseudo-Marginal MCMC for Inference of the Stable Distribution Parameters

In this chapter, we propose a novel PM sampling scheme for inference of the parameters of the scalar, arbitrarily skewed, α-stable distribution, based on the LV model that stems from the CMS generative scheme. The only restriction on the parameters that we assume is that $\alpha \neq 1$, an assumption commonly made in the α-stable literature, as the mathematical expressions involved are not well defined in this case, requiring it to be treated separately. The CMS model was presented in Section 2.4.1, the general framework for PM samplers was introduced in Section 3.2.3, while in Section 3.3.1 we summarized the GS proposed in the literature by Buckle [1995].

In the GS, sampling the LVs involves numerical operations that make the sampler approximate. Moreover, the LVs appear to be highly correlated with the parameters, and a re-parametrization is suggested to improve the mixing of the parameter chains, which requires further numerical approximations. On the contrary, while the same LVs are drawn in the PM sampler, these are used to form an unbiased estimator of the likelihood based on the IS scheme, whose exactness is not affected by the numerical approximation introduced. Moreover, in the PM sampler the LVs are used to mimic an ideal marginal scheme: in this way, in the PM scheme, the correlation between the sampled variables is naturally reduced, and it is possible to target the original full-conditional distributions of the parameters, without re-parametrizing the LVs, thus avoiding the second source of approximation of the GS. However, the PM scheme is more computationally intensive than the GS, because a set of $M > 1$ LVs per stable data point need to be sampled at each iteration, in order to compute

the likelihood estimate; on the contrary, the GS only requires one LV for each stable data point.

The structure of this chapter is as follows: we first present some analytic considerations on the nature of the conditional distribution of the LVs. These considerations justify why Buckle [1995] resorted to an adaptive mechanism for proposing the LVs in the GS scheme, as well as its adaptation that we propose for the PM scheme. In particular, we generalize the IS presented in Algorithm 5, in order to exploit also the samples that are used to adaptively build the proposal. Furthermore, we show that adaptive proposal distributions for the LVs generate likelihood estimates with lower variance than those induced by standard (non-adaptive) choices. Finally, we compare our proposed PM sampler with the GS by Buckle [1995].

Observe that, for simplicity, in this chapter we work with standardized stable data, aiming at inferring $\boldsymbol{\theta} = (\alpha, \beta)$, which are the most characteristic parameters of the α-stable distribution. A generalization to the four parameters $(\alpha, \beta, \sigma, \mu)$ is conceptually straightforward, but likely to increase the variance of the estimates, as well as the computational time.

4.1 Properties of the Conditional Distribution of the LVs

Recall from Section 2.4.1 that the PDF of a standard stable RV $Z \sim \mathcal{S}_\alpha(1, \beta, 0)$[1] is the marginal of the bivariate distribution $p_{(Z,Y)}(z,\, y|\alpha,\, \beta)$ defined in (2.21), and that we report for easy reference:

$$p_{(Z,Y)}(z,\, y|\alpha,\, \beta) = \frac{\alpha}{|\alpha - 1|} \exp\left\{ -\left| \frac{z}{t_{\alpha,\beta}(y)} \right|^{\alpha/(\alpha-1)} \right\} \left| \frac{z}{t_{\alpha,\beta}(y)} \right|^{\alpha/(\alpha-1)} \frac{1}{|z|},$$

where the function $t_{\alpha,\beta}$, defined for $y \in (-1/2, 1/2)$, is given in (2.8),

$$t_{\alpha,\beta}(y) := \frac{\sin(\pi\alpha y + \eta_{\alpha,\beta})}{(\cos(\pi y))^{1/\alpha}} \left(\cos((\alpha - 1)\pi y + \eta_{\alpha,\beta}) \right)^{\frac{1-\alpha}{\alpha}}, \tag{4.1}$$

and $\eta_{\alpha,\beta}$ is a constant (2.9).

If $\boldsymbol{z} := [z_1, \ldots, z_N]$ is the standard stable dataset, and $\boldsymbol{y} := [y_1, \ldots, y_N]$ is the corresponding vector of LVs, one per stable data point, in the CMS LV model, then,

[1] In this Chapter we refer to the parametrization (2.4), but we use a simplified notation, omitting the subscripts on the skeweness and scale parameters.

by conditional independence, the posterior full conditional distribution of the LVs is

$$p(\boldsymbol{y}|\boldsymbol{z},\alpha,\beta) = \prod_{i=1}^{N} p(y_i|z_i,\alpha,\beta) \propto p(\boldsymbol{z},\boldsymbol{y}|\alpha,\beta) = \prod_{n=1}^{N} p_{(Z,Y)}(z_n,y_n|\alpha,\beta), \quad (4.2)$$

or, equivalently, $p(y_n|z_n,\alpha,\beta) \propto p_{(Z,Y)}(z_n,y_n|\alpha,\beta)$. The distribution $p(y_n|z_n,\alpha,\beta)$ plays a key role in both the GS and the PM schemes: in the first case, it has to be sampled from, while, in the second case, it is the target of the IS step used to compute an unbiased estimate of the likelihood.

4.1.1 Properties of $t_{\alpha,\beta}$

Given the proportionality relation (4.2), the function $t_{\alpha,\beta}$, that appears in the expression of $p_{(Z,Y)}$, is essential to determine the properties of the posterior full conditional distribution of the latent variables. Buckle [1994] provides an analytic study, that can be summarized as follows:

(i) $t_{\alpha,\beta}(y) = -t_{\alpha,-\beta}(-y)$;

(ii) $t_{\alpha,\beta}(y) = 0$ iff $y = l_{\alpha,\beta}$, where $l_{\alpha,\beta}$ is defined in (2.20);

(iii) $t_{\alpha,\beta} : (-1/2, 1/2) \mapsto (-\infty, \infty)$, is a continuous and strictly monotonic function, and, for $\alpha \neq 2$, $t_{\alpha,\beta}(\pm 1/2) = \pm\infty$;

(iv) the derivative of $t_{\alpha,\beta}$ wrt its argument can be expressed as [2]

$$t'_{\alpha,\beta}(y) = \alpha\pi \left[\cos \pi y\right]^{(\alpha-1)/\alpha} \left[\cos(\pi(\alpha-1)y + \eta_{\alpha,\beta})\right]^{1/\alpha} \times \\ \times \left[1 + \frac{1}{\alpha^2}\left(\frac{\sin \pi y}{\cos \pi y} - (\alpha-1)\frac{\sin[\pi(\alpha-1)y + \eta_{\alpha,\beta}]}{\cos[\pi(\alpha-1)y + \eta_{\alpha,\beta}]}\right)^2\right]; \quad (4.3)$$

(v) When $\alpha = 2$, we have $l_{2,\beta} = 0$, $t_{2,\beta}(y) = 2\sin(\pi y)$ and $t'_{2,\beta}(y) = 2\pi\cos(\pi y)$; observe that the image of $t_{2,\beta}(y)$ is bounded in this case to $t_{2,\beta}(y) \in [-2, 2]$, and that $t'_{2,\beta}(\pm 1/2) = 0$.

Figure 4.1 shows some instances of the function $t_{\alpha,\beta}(y)$ for a number of choices of the parameters α and β. A remarkable feature is that, despite being monotonically increasing, the function cannot be analytically inverted, an operation required throughout this chapter. Moreover, the function is 'almost flat' for small values of α, meaning that it is more difficult to numerically invert it in that parameter regime.

[2] A number of alternative expressions of $t'_{\alpha,\beta}$ are given in Buckle [1994] and Buckle [1995]. Here we use the version reported in the first reference, while that in the second one appears to be incorrect.

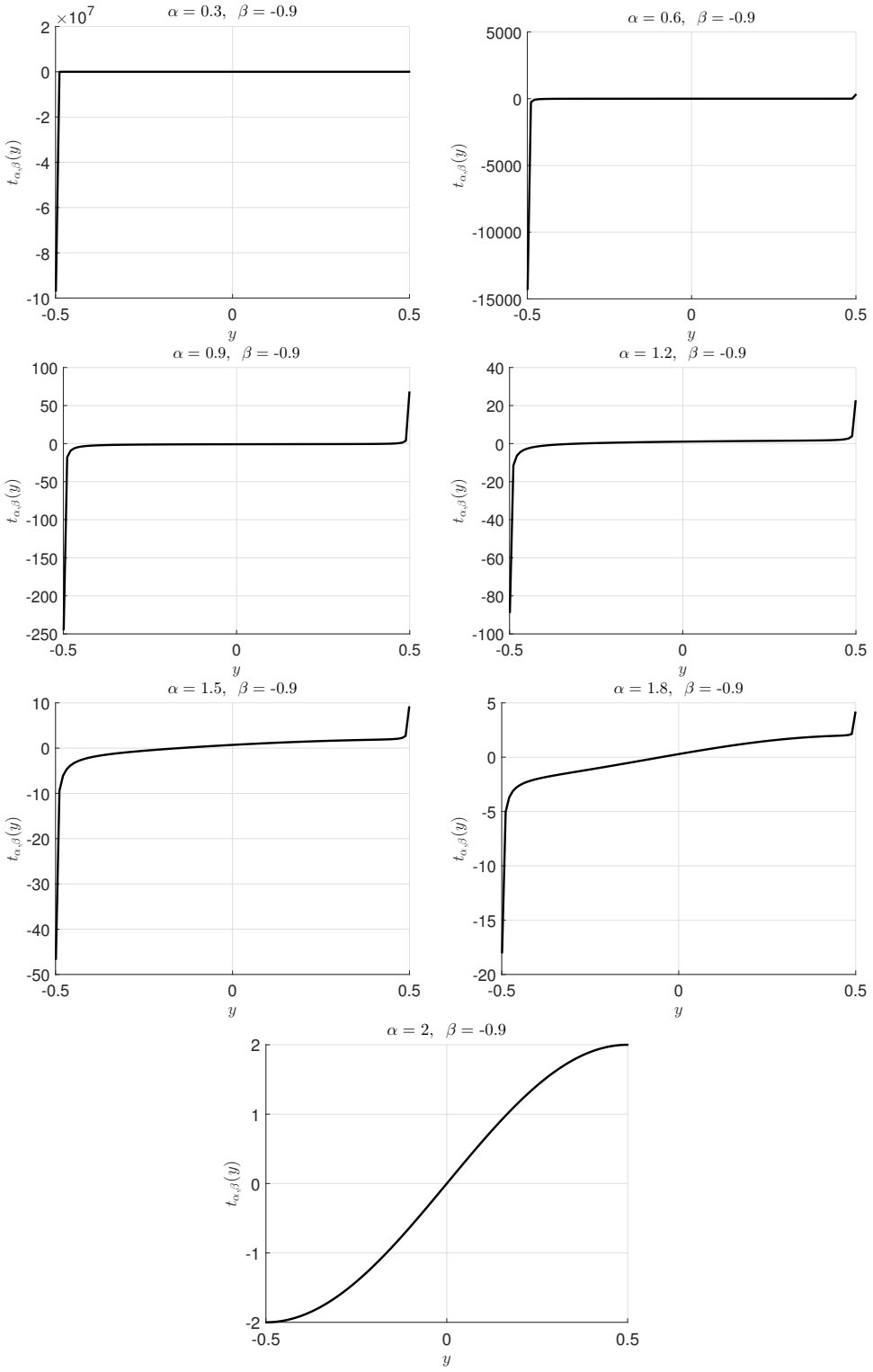


Figure 4.1 Function $t_{\alpha,\beta}(y)$, varying α, and with $\beta = -0.9$.

4.1.2 Properties of $p(y_n|z_n, \alpha, \beta)$

From the properties of $t_{\alpha,\beta}$ and from the definition of the support of $p_{(Z,Y)}(z_n, y_n|\alpha, \beta)$, it follows that[3]

$$\operatorname{sgn}(z_n) = \operatorname{sgn}(t_{\alpha,\beta}(y_n)), \tag{4.4}$$

and it is thus possible to remove the absolute values in (2.21) every time that the ratio $z/t_{\alpha,\beta}(y)$ appears. Hence the posterior full-conditional distribution of each LV is proportional to[4]

$$\begin{aligned} p(y_n|z_n, \alpha, \beta) \propto p_{(Z,Y)}(z_n, y_n|\alpha, \beta) &\propto \exp\left\{-\left|\frac{z_n}{t_{\alpha,\beta}(y_n)}\right|^{\alpha/(\alpha-1)} + 1\right\}\left|\frac{z_n}{t_{\alpha,\beta}(y_n)}\right|^{\alpha/(\alpha-1)} \\ &= \exp\left\{-\left(\frac{t_{\alpha,\beta}(y_n)}{z_n}\right)^{\alpha/(1-\alpha)} + 1\right\}\left(\frac{t_{\alpha,\beta}(y_n)}{z_n}\right)^{\alpha/(1-\alpha)} \end{aligned} \tag{4.5}$$

defined for $y_n \in \mathcal{Y}_n$, as in (2.22). To simplify the notation, we define

$$t(y) := \left(\frac{t_{\alpha,\beta}(y)}{z}\right)^{\alpha/(1-\alpha)},$$

with first derivative

$$t'(y) = \frac{\alpha}{1-\alpha}\left(\frac{t_{\alpha,\beta}(y)}{z}\right)^{(2\alpha-1)/(1-\alpha)} \frac{1}{z} t'_{\alpha\beta}(y), \tag{4.6}$$

and second derivative $t''(y)$.

Then the LV full-conditional (4.5) is of the form $p(y_n|z_n\alpha, \beta) \propto g(y_n)$, with

$$g(y) := \exp(-t(y) + 1)t(y),$$

and

$$g'(y) = t'(y)\exp(-t(y) + 1)(-t(y) + 1), \tag{4.7}$$

$$g''(y) = t''(y)[\exp(-t(y) + 1)(-t(y) + 1)] - (t'(y))^2 \exp(-t(y) + 1)[-t(y) + 2]. \tag{4.8}$$

Thus the following properties hold for $p(y_n|z_n, \alpha, \beta)$, some of which are summarized by Buckle [1995]:

[3] We handle separately the case $\alpha = 2$, in which the image of $t_{\alpha,\beta}$ is bounded.

[4] Through a positive multiplicative constant. We arbitrarily introduce or remove multiplicative constants in a way that makes it easier to determine the properties of $p(y_n|z_n, \alpha, \beta)$.

(i) $p(\pm 1/2|z_n, \alpha, \beta) = 0$, for $\alpha \neq 2$, and $p(l_{\alpha,\beta}|z_n, \alpha, \beta) = 0$, $\forall \alpha$, thanks to the properties 2 and 3 of $t_{\alpha,\beta}$ above;

(ii) considering (4.7), it is possible to see that $p(y_n|z_n\alpha, \beta)$ has a unique stationary point. In fact, $g'(y_n) = 0$ if either $t(y_n) = 1$ or $t'(y_n) = 0$.
The first condition is achieved only for the value y_n^* such that $t_{\alpha,\beta}(y_n^*) = z_n$, namely $y_n^* = t_{\alpha,\beta}^{-1}(z_n)$; given that $z_n \in \mathbb{R}$, the last expression is appropriate if $\alpha \neq 2$ or, if $\alpha = 2$ for those values of $z_n \in [-2, 2]$.
On the other hand, from (4.6)[5] the condition $t'(y_n) = 0$, is equivalent to $t'_{\alpha,\beta}(y_n) = 0$, and this is possible only for $\alpha = 2$, in which case $t'_{2,\beta}(y_n) = 2\pi \cos(\pi y_n) = 0$ only if $y_n = y_n^* = \pm 1/2$, with $y_n^* = 1/2 \in \mathcal{Y}_n$ if $z_n > 0$ and $y_n^* = -1/2 \in \mathcal{Y}_n$ if $z_n < 0$.

Thus, the unique stationary point y_n^* is

$$y_n^* := \begin{cases} t_{\alpha,\beta}^{-1}(z_n) & \text{if } \alpha \in (0,1) \cup (1,2) \vee \alpha = 2, |z_n| < 2, \\ -\frac{1}{2} = t_{\alpha,\beta}^{-1}(-2) & \text{if } \alpha = 2, z_n \leq -2, \\ \frac{1}{2} = t_{\alpha,\beta}^{-1}(2) & \text{if } \alpha = 2, z \geq 2. \end{cases} \tag{4.9}$$

The inversion of $t_{\alpha,\beta}$ necessary to compute y_n^* cannot be done in closed form. As suggested in Buckle [1995], we solve this problem by applying a safeguarded Newton method, see Gill et al. [1981]. This consists of a bisection step every time the Newton search locates the approximated zero of the function $h(y_n) := t_{\alpha,\beta}(y_n) - z_n$ out of $\mathcal{Y}_n$, in which we know the zero should lie, thanks to (4.4). The more $\mathcal{Y}_n$ is concentrated around $1/2$ or $-1/2$, where $t_{\alpha,\beta}$ has vertical asymptotes, the more iterations are required to find a good solution (namely to guarantee that the distance of the approximate zero of h to the true one is lower than a given tolerance value);

Observe also that, when evaluated in y_n^*, expression (4.5) is equal to one,

$$p(y_n^*|z_n, \alpha, \beta) = p_n^* = 1 \tag{4.10}$$

but this is only proportional to the value of the full-conditional distribution.

(iii) using (4.8), when evaluated at the stationary point, the second derivative of the LV full-conditional becomes

$$p''(y_n^*|z_n, \alpha, \beta) \propto -\left[t'(y_n)\right]^2 < 0, \tag{4.11}$$

[5] The cases $y_n = \pm 1/2$ and $y_n = l_{\alpha,\beta}$, corresponding to the asymptotes and zero of $t_{\alpha,\beta}$ have been considered separately above.

with $t'(y_n)$ as in (4.6).This proves that y_n^* is a maximum.

The main implication of the above properties is that $p(y_n|z_n, \alpha, \beta)$ is a *unimodal* distribution with *compact support* $\mathcal{Y}_n$. These are desirable features, because, in theory, they imply that it is easy to sample from $p(y_n|z_n, \alpha, \beta)$, a task required both in the GS and the PM schemes for estimating the posterior distribution of the parameters α and β. However, the 'peakiness' of $p(y_n|z_n, \alpha, \beta)$ heavily depends on the value of the variables that it is conditioned on: the realization of the stable data point $|z_n|$, and the range of the parameters α and β. In particular, the LV full-conditional distribution appears to be very peaked for very small or large values of $|z_n|$, a situation that induces high rejection rates or slow mixing in standard samplers. Figure 4.2 shows instances of the joint PDF $p_{(Z,Y)}(z, y|\alpha, \beta)$ for a range of parameters α and β, while Figure 4.3 shows $p(y_n|z_n, \alpha, \beta)$, for a number of fixed values of z_n for each choice of the parameters α and β. Observe that in the second graph, the distribution corresponds to 'slices' of $p_{(Z,Y)}(z, y|\alpha, \beta)$, and it has not been scaled to normlize the mode equal to 1.

In the following sections we detail the adaptive RS scheme that was proposed by Buckle [1995] to sample the LV for the GS scheme. We take inspiration from such adaptive RS for proposing the LV in the IS step for estimating the likelihood in the PM scheme, and we compare the variance of the likelihood estimate with that obtained with more standard proposals.

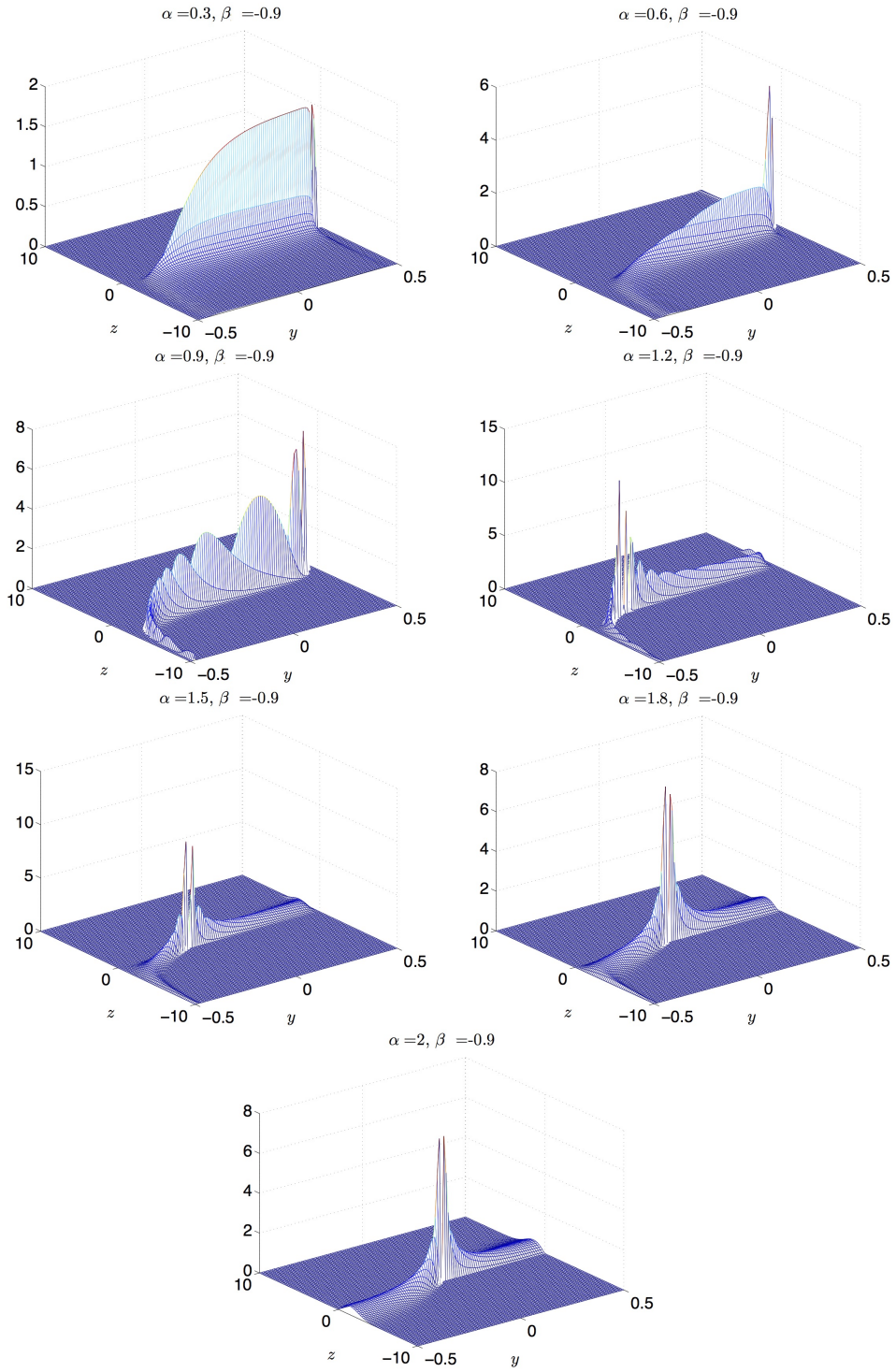


Figure 4.2 Joint PDF $p_{(Z,Y)}(z, y|\alpha, \beta)$ of the standard stable RV Z and the latent variable Y, varying α, with $\beta = -0.9$.

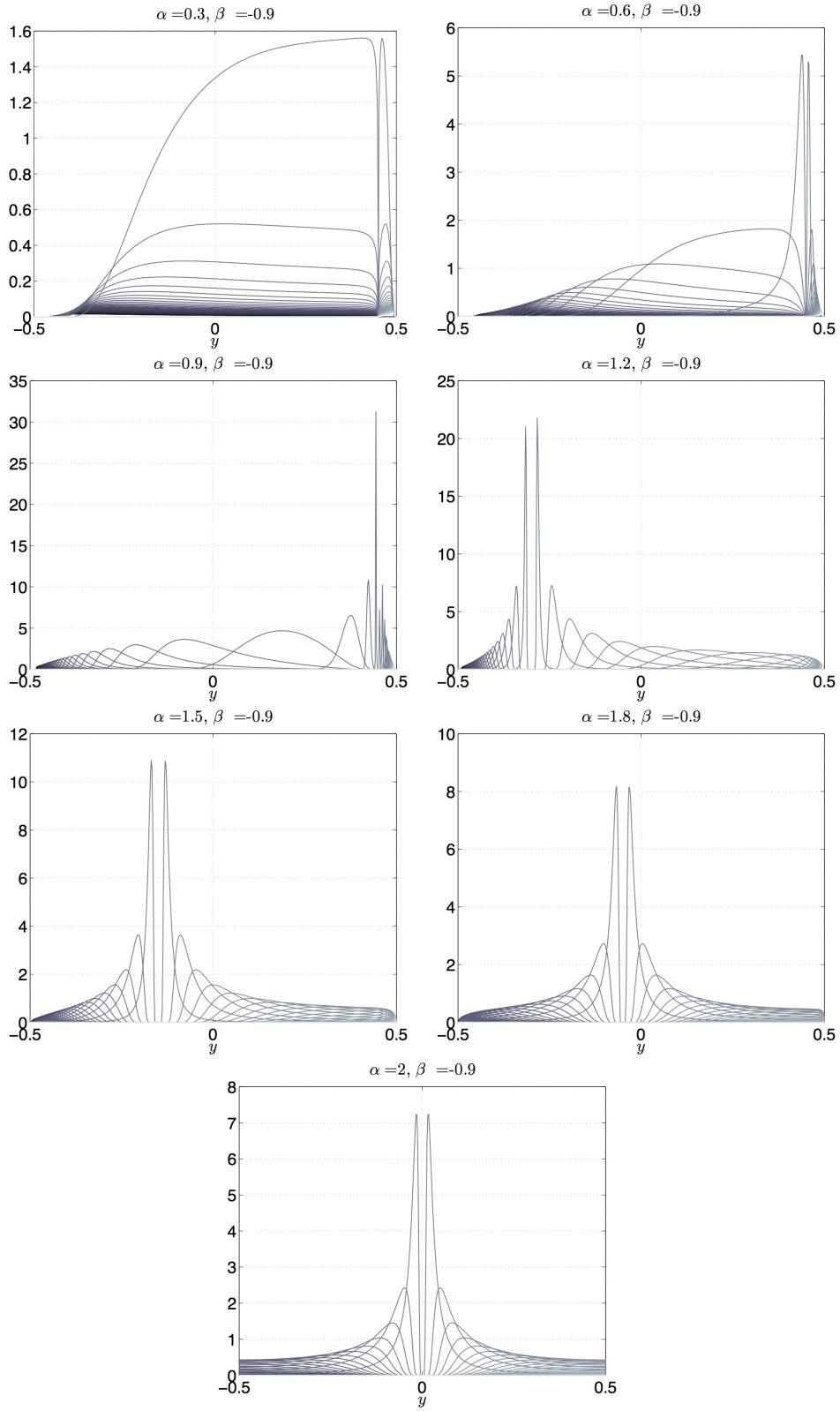


Figure 4.3 Posterior full-conditional distribution of the LV Y, $p(y_n|z_n, \alpha, \beta)$, for 100 different values of $z_n \in (-10, 10)$, for each choice of the parameters α and β.

4.2 Sampling the LVs in the GS scheme

Let us denote with y_n^* the unique maximum point (4.9) of $p(y_n|z_n, \alpha, \beta)$, with $p_n^* := p(y_n^*|z_n, \alpha, \beta)$ its maximum value, and with $\mathcal{Y}_n = [-1/2, l_{\alpha,\beta}]$, if $z_n < 0$ or $\mathcal{Y}_n = [l_{\alpha,\beta}, 1/2]$, if $z_n > 0$, its compact support. Theoretically, unimodality on a compact support makes RS with a uniform proposal on $\mathcal{Y}_n$, with PDF proportional to[6]

$$q_1(y_n|z_n, \alpha, \beta) := p_n^* \mathbb{1}_{\mathcal{Y}_n}(y_n), \tag{4.12}$$

a suitable algorithm for sampling from $p(y_n|z_n, \alpha, \beta)$. According to Buckle [1995], this is proved to be the case for most of the data sets examined.

However, given that, for very large or small values of $|z_n|$, $p(y_n|z_n, \alpha, \beta)$ is fairly spiked around the mode, the RS algorithm suffers of poor acceptance rates, when such values are present in the dataset. Buckle [1995] proposes to use an adaptive RS mechanism that makes uses of the rejected points to draw samples from the LV full-conditional distribution. Assume that a point Y_1 uniformly sampled from $q_1(y_n|z_n, \alpha, \beta)$ has been rejected. We know that, if $t_{\alpha,\beta}(Y_1) < z_n$, then Y_1 lies to the left of the mode of $p(y_n|z_n, \alpha, \beta)$, and if $t_{\alpha,\beta}(Y_1) > z_n$, then Y_1 lies to the right of its mode. Thus, the new rejection envelope becomes the histogram, i.e. the piece-wise constant density function, with heights p_n^* and $p(Y_1|z_n, \alpha, \beta)$. For example, if $z_n < 0$ the two-piece constant envelope function is[7]

$$q_2(y_n|z_n, \alpha, \beta, Y_1) := \begin{cases} p_1 \mathbb{1}_{A_1}(y_n) + p_2 \mathbb{1}_{A_2}(y_n) & \text{if } t_{\alpha,\beta}(Y_1) < z_n, \\ p_2 \mathbb{1}_{A_1}(y_n) + p_1 \mathbb{1}_{A_2}(y_n) & \text{if } t_{\alpha,\beta}(Y_1) > z_n, \end{cases}$$

where $p_1 = p(Y_1|z_n, \alpha, \beta)$, $A_1 = (-1/2, Y_1)$, and $p_2 = p_n^*$, $A_2 = (Y_1, l_{\alpha,\beta})$. The procedure is repeated by sampling the next point Y_2 uniformly over one of the two segments, where the segment is selected from a categorical distribution with parameters (π_1, π_2) given by the normalized areas of the two segments. If the sample Y_2 is accepted in the RS, then this is an exact draw from the full-conditional distribution of the LV; if Y_2 is rejected, then it is added to the histogram envelope, that gradually converges to the true density and hence gains a considerable efficiency increase. According to Buckle [1995], rarely more than 6 (non i.i.d.) candidate points

[6] Recall that $\mathbb{1}_{\mathcal{Y}_n}(y_n)$ is the indicator function equal to 1 if $y_n \in \mathcal{Y}_n$, and 0 otherwise.
[7] We stress here the dependence of the proposal distribution on the rejected LVs.

are required before acceptance, implying that the last envelope proposal is

$$q_G(y_n|z_n,\alpha,\beta,Y_{1:G-1}) = \sum_{j=1}^{G} p_j \mathbb{1}_{A_j}(y_n), \tag{4.13}$$

with p_j and A_j appropriately defined, $Y_{1:j-1} := [Y_1,\ldots,Y_{G-1}]$ a vector of the rejected values of $G-1$ LVs before acceptance, and G a random number typically smaller than 7.

Given that the LVs are conditionally independent, it is sufficient to repeat this procedure N times, one for each LV, to obtain a sample from $p(\boldsymbol{y}|\boldsymbol{z},\alpha,\beta)$ in (4.2) at the k-th GS step, as required by (3.31a) in Section 3.3.1.

4.3 Sampling the Likelihood in the PM scheme

The numerical inversion of the function $t_{\alpha,\beta}$ performed in order to obtain (y_n^*,p_n^*) is a source of approximation in the GS, because the RS acceptance criteria of the kind (3.8) are evaluated in an approximate way. Moreover, in the GS scheme by Buckle [1995], the LVs appear to be highly correlated with the parameters, making it necessary to re-parametrize the full-conditional distribution of the parameters to allow for chain mixing, but this introduces further approximation, as detailed in Section 4.4. To overcome these issues, our novel contribution is the implementation of an efficient PM scheme that, instead of the LVs, samples an unbiased estimate of the likelihood. As explained in Section 3.2.3, the successful implementation of the PM scheme depends on the ability of producing likelihood estimators ζ that have small variance.

Given the integral representation of the likelihood (2.23), we consider IS based estimators, as in equation (3.21). The IS step is defined by Algorithm 5, once $q(y_n|z_n,\alpha,\beta)$, the proposal distribution for each LV given the parameters and the standardized stable data point, is set. According to the considerations in Section 3.2.3, the likelihood estimator has small variance when $q(y_n|z_n,\alpha,\beta)$ is 'close' to $p(y_n|z_n,\alpha,\beta)$, the full-conditional distribution of the LVs. Observe that the IS estimate of the likelihood $\hat{p}(\boldsymbol{z}|\alpha,\beta)$ is the product of terms that are typically either very small or very large, making the computation of the likelihood not numerically stable. It is thus advised to work with robust expressions of the log-likelihood, and

then exponentiate them. In particular, considering (3.21) and (3.22), we can write

$$\begin{aligned}\log \hat{p}(z_n|\alpha,\beta) &= \log\left(\sum_{j=1}^{M} W_j\right) - \log M \\ &= \log c_n + \log\left(\sum_{j=1}^{M} W_j c_n^{-1}\right) - \log M, \end{aligned} \tag{4.14}$$

for any arbitrary constant $c_n > 0$, possibly different for each stable data point z_n. Then, decomposing

$$q(y_n|z_n,\alpha,\beta) = \frac{\tilde{q}(y_n|z_n,\alpha,\beta)}{Q(z_n,\alpha,\beta)}, \tag{4.15}$$

where $\tilde{q}$ is the unnormalized proposal PDF, and Q is its normalizing constant, in order to scale to 1 the maximum of the terms $W_j c_n^{-1}$, in (4.14) we take $c_n := \max_j W_j$, so that

$$\begin{aligned}\log c_n = \log \max_j W_j &= \max_j \log W_j \\ &= \max_j \Big(\log p(z_n, Y_j|\alpha,\beta) - \log \tilde{q}(Y_j|z_n,\alpha,\beta) + \log Q(z_n,\alpha,\beta)\Big) \\ &= \max_j \log\Big(p(z_n, Y_j|\alpha,\beta)/\tilde{q}(Y_j|z_n,\alpha,\beta)\Big) + \log Q(z_n,\alpha,\beta) \\ &= \log \tilde{c}_n + \log Q(z_n,\alpha,\beta),\end{aligned}$$

where we have defined $\tilde{c}_n := \max_j(p(z_n, Y_j|\alpha,\beta)/\tilde{q}(Y_j|z_n,\alpha,\beta))$. Moreover, using the same computations:

$$\begin{aligned}W_j c_n^{-1} &= \exp\Big(\log W_j - \log c_n\Big) \\ &= \exp\Big(\log p(z_n, Y_j|\alpha,\beta) - \log \tilde{q}(Y_j|z_n,\alpha,\beta) - \log \tilde{c}_n\Big),\end{aligned}$$

and thus (4.14) becomes

$$\begin{aligned}\log \hat{p}(z_n|\alpha,\beta) = &\left(\log \tilde{c}_n + \log Q(z_n,\alpha,\beta)\right) - \log M \\ &+ \log\left(\sum_{j=1}^{M} \exp\left(\log p(z_n, Y_j|\alpha,\beta) - \log \tilde{q}(Y_j|z_n,\alpha,\beta) - \log \tilde{c}_n\right)\right).\end{aligned} \tag{4.16}$$

Finally, the likelihood estimator is computed as

$$\hat{p}(\boldsymbol{z}|\alpha,\beta) = \exp\left(\sum_{n=1}^{N} \log \hat{p}(z_n|\alpha,\beta)\right) = \exp(\hat{\ell}(\boldsymbol{z}|\alpha,\beta)), \tag{4.17}$$

where we have introduced the notation $\hat{\ell}(\boldsymbol{z}|\alpha,\beta) := \log \hat{p}(\boldsymbol{z}|\alpha,\beta)$ to indicate the logarithm of the likelihood estimator. It is clear that expression (4.16) is determined once $q(y_n|z_n,\alpha,\beta)$ is set, i.e. once we have defined the mechanism for sampling the LVs $\{Y_j\}_{j=1}^M$ in the IS scheme, and we are able to evaluate the unnormalized proposal density $\tilde{q}$ and its normalizing constant Q, as expected. In the following, we examine a number of different choices of $q(y_n|z_n,\alpha,\beta)$, that exploit the 'unimodality on a compact support' property of $p(y_n|z_n,\alpha,\beta)$, proved in Section 4.1. We then compare the variance of the likelihood estimators thus obtained.

In the PM scheme the likelihood estimator is unbiased, namely $\mathbb{E}[\hat{p}(\boldsymbol{z}|\alpha,\beta)] = p(\boldsymbol{z}|\alpha,\beta)$,[8] and its variance is

$$\mathrm{Var}[\hat{p}(\boldsymbol{z}|\alpha,\beta)] = \mathbb{E}[\hat{p}^2(\boldsymbol{z}|\alpha,\beta)] - p^2(\boldsymbol{z}|\alpha,\beta). \tag{4.18}$$

On the other hand, the logarithm of the likelihood estimator is typically a biased estimator for the log-likelihood. In fact, by Jensen's inequality [Feller, 1966, p. 153],

$$\mathbb{E}[\hat{\ell}(\boldsymbol{z}|\alpha,\beta)] = \mathbb{E}[\log \hat{p}(\boldsymbol{z}|\alpha,\beta)] \leq \log \mathbb{E}[\hat{p}(\boldsymbol{z}|\alpha,\beta)] = \log p(\boldsymbol{z}|\alpha,\beta).$$

The larger the variance of $\hat{p}(\boldsymbol{z}|\alpha,\beta)$ (4.18), the larger the bias of its logarithm $\mathbb{E}[\hat{\ell}(\boldsymbol{z}|\alpha,\beta)] - \log p(\boldsymbol{z}|\alpha,\beta)$. In fact, again by Jensen's inequality $\mathbb{E}[\hat{p}^2(\boldsymbol{z}|\alpha,\beta)] \geq p^2(\boldsymbol{z}|\alpha,\beta)$, with concavity of the function $\hat{p}^2$ opposite to that of $\log \hat{p}$. Thus, in the following, we will consider logarithms of the likelihood estimators, and we prefer those that have smaller bias. Of course, we cannot compute this bias exactly, because the likelihood $p(\boldsymbol{z}|\alpha,\beta)$ cannot be computed in first instance; however, it is possible to have an intuition of the bias by examining boxplots of realizations of $\hat{\ell}(\boldsymbol{z}|\alpha,\beta)$, corresponding to different draws of the LVs $\{Y_j\}_{j=1}^M$.

4.3.1 Piece-wise Constant and Linear Adaptive Proposals

The first method that we use[9] to draw $\{Y_j\}_{j=1}^M$ is inspired by the adaptive mechanism for sampling the LVs by Buckle [1995], presented in Section 4.2. In particular, the unnormalized proposal distribution $\tilde{q}$ can be obtained by performing a number of adaptations of the piece-wise constant envelope function on $p(y_n|z_n,\alpha,\beta)$. This is achieved by sampling a first group of $G-1$ (not identically distributed) LVs $\{Y_j\}_{j=1}^{G-1}$,

[8] Recall that the expectations are taken wrt the proposal distribution of the LVs $q(\boldsymbol{y}|\boldsymbol{z},\alpha,\beta) = \prod_{n=1}^N q(y_n|z_n,\alpha,\beta)$.

[9] This was published in Riabiz et al. [2015].

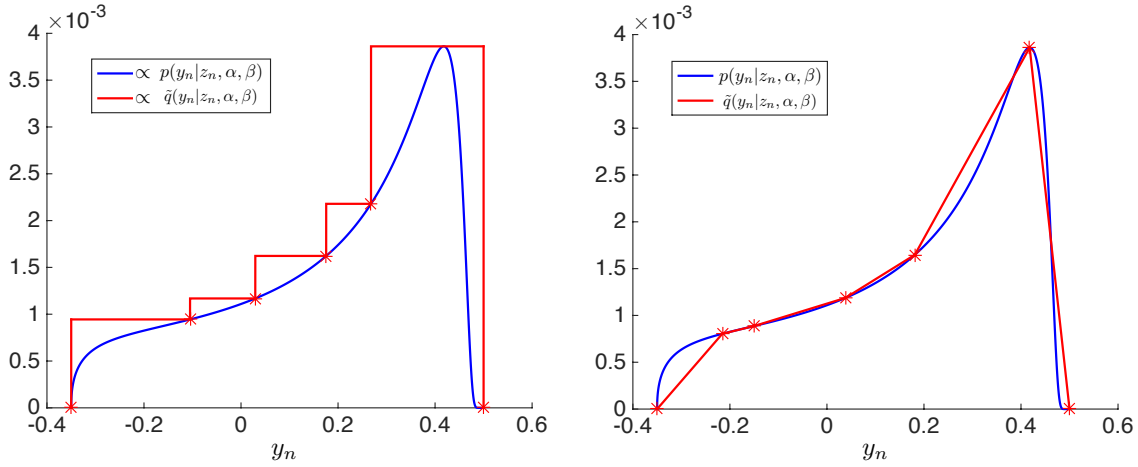


Figure 4.4 Adaptive proposal distributions $\tilde{q}(y_n|z_n, \alpha, \beta)$ in the PM sampler, with $G = 5$ adaptations, and targeting $p(y_n|z_n, \alpha, \beta)$, with $z_n = 40.83$, $\alpha = 0.3$ and $\beta = 0.7$. Left: piece-wise constant envelope function. Right: piece-wise linear interpolating function. The red stars correspond to the adaptively sampled points, the extremes of $\mathcal{Y}_n$, and the maximum point (y_n^*, p_n^*) used to initialize the piece-wise linear proposal.

starting from the uniform envelope (4.12), and obtaining[10]

$$\tilde{q}(y_n|z_n, \alpha, \beta) := q_G(y_n|z_n, \alpha, \beta, Y_{1:G-1}), \tag{4.19}$$

where q_G is defined as in (4.13), but now the number of adaptations $G-1$ is fixed *a priori*; unlike the RS by Buckle [1995], no accept-reject step is performed, but all the sampled points $\{Y_j\}_{j=1}^G$ are used to adapt the envelope function. The normalizing constant $Q(z_n, \alpha, \beta, Y_{1:G-1})$ is easily computed as sum of G rectangular areas, one of each of the disjoint partitions A_j, whose union is $\mathcal{Y}$. Thus, when normalized, (4.19) forms the proposal distribution $q(y_n|z_n, \alpha, \beta) = q(y_n|z_n, \alpha, \beta, Y_{1:G-1})$, from which the M i.i.d. samples $\{Y_j\}_{j=1}^M$ in Algorithm 5 are drawn.

Observe that, when no adaptation is done ($G - 1 = 0$), $q_1(y_n|z_n, \alpha, \beta)$ is simply the uniform proposal. Here we assume that $G \geq 2$, and that at least one adaptation of the envelope function is done, but we use the uniform proposal for comparisons in Section 4.5.1.

The idea of adaptive proposals can be extended by choosing $\tilde{q}(y_n|z_n, \alpha, \beta)$ to be a piece-wise linear interpolating function on $p(y_n|z_n, \alpha, \beta)$, that is intuitively likely to resemble more the target, leading to smaller variance of the likelihood estimator. The first difference wrt the piece-wise constant proposal is in the initialization, through a triangular function that interpolates $p(y_n|z_n, \alpha, \beta)$ at the extremes of $\mathcal{Y}_n$ and in the maximum point (y_n^*, p_n^*). Subsequently, $G-1$ points are sampled from the piece-wise

[10]The notation below underlines the adaptive nature of the proposal distribution.

Algorithm 7 Adaptive proposals for LV sampling in the PM scheme.

1: Set $q_1(y_n|z_n,\alpha,\beta) = \begin{cases} p_n^* \mathbb{1}_{\mathcal{Y}_n}(y_n) \text{ (for PCE)} \\ \text{triangle with maximum in } (y_n^*, p_n^*) \text{ (for PLI)} \end{cases}$
2: **for** $j = 1, \ldots, G-1$ **do**
3: sample $Y_j \sim q_j(y_n|z_n,\alpha,\beta)$:
4: choose an interval A_j (discrete distribution)
5: and sample $\begin{cases} \text{uniformly on } A_j \text{ (for PCE)} \\ \text{invert a quadratic CDF on } A_j \text{ (for PLI)} \end{cases}$
6: add a constant (for PCE) or oblique (for PLI) segment on A_j, to get $q_{j+1}(y_n|z_n,\alpha,\beta)$
7: **return** $q_G(y_n|z_n,\alpha,\beta)$

linear PDF, by first selecting the segment A_j to sample from,[11] followed by inversion of the quadratic CDF, see Devroye [1986], corresponding to the linear PDF on the chosen interval. The normalizing constant $Q(z_n,\alpha,\beta)$ is easily computed as sum of trapezoidal (or triangular) areas, one for each A_j. The extra complexity added wrt the piece-wise constant envelope method is thus in solving $G-1$ quadratic equations, one for each of the points that are adaptively sampled; this can be considered a constant factor that does not increase significantly the overall computational cost.

Figure 4.4 shows an example of the two procedures, while Algorithm 7 gives a summary of both (in the Algorithm, PCE denotes the piece-wise constant envelope proposal, while PLI is the piece-wise linear interpolating proposal).

Remark 4. The approximate location of y_n^*, the maximum point of $p(y_n|z_n,\alpha,\beta)$, does not affect the exactness of the IS step in the PM scheme. In fact, in the IS the (unnormalized) proposal does not need to be above the target, and it is thus not necessary to exactly find (y_n^*, p_n^*). Notice also that the bias introduced in the GS by the numerical procedure appears not to have a significant effect on the simulation results; see the posterior distributions obtained with the two schemes in Section 4.5.2.

4.3.2 Cost of the LV Adaptive Proposals

In the adaptive proposal described above, $G-1$ denotes the number of adaptations of each of the LV proposal distributions $q(y_n|z_n,\alpha,\beta)$, and M the number of i.i.d. draws from it. We can identify two sources of computational cost, in this geometrically-informed adaptive method for sampling the likelihood estimator in the IS-based PM scheme:

[11] As for the piece-wise constant envelope, this is done by sampling from a discrete distribution that has probabilities corresponding to the normalized areas of each triangular section.

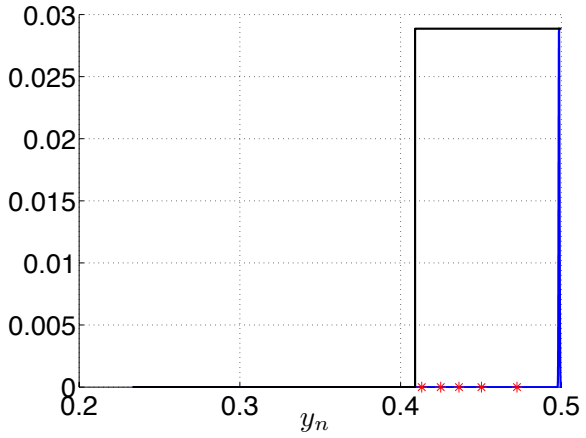


Figure 4.5 Case of peaked LV conditional distribution (blue), with $z_n = 76.5$, $\alpha = 1.2$ and $\beta = 0.7$; the piece-wise constant envelope proposal for the LV (black) has $G = 5$ levels, and only $M = 5$ points are sampled from it (red stars).

- the cost of performing $G - 1$ adaptations of $q(y_n|z_n, \alpha, \beta)$; for each adaptation this involves sampling from a discrete distribution, followed by sampling from a distribution with either constant or linear PDF, and modifying the variables saved in the computer memory (typically the extremes of the segments A_j and a scalar, indicating either a height or a slope value) to pass from $q_j(y_n|z_n, \alpha, \beta)$ to $q_{j+1}(y_n|z_n, \alpha, \beta)$.
- the cost of evaluating the function $p(y_n|z_n, \alpha, \beta)$, or, alternatively, $p(y_n, z_n|\alpha, \beta)$. Observe that this is required both during the $G - 1$ adaptations of the LV proposal, and when computing the weights (3.22) for the M draws from $q(y_n|z_n, \alpha, \beta)$. Thus the total number of evaluations of the LV full conditional is

$$L := G + M, \tag{4.20}$$

  given that one evaluation is needed also to initialize the proposals at uniform height p_n^*.[12]

In the following we analyse strategies to reduce both of these costs.

Trade-off between G and M

Reducing the adaptation cost corresponds to choosing a low value of G. This is a good strategy if the target $p(y_n|z_n, \alpha, \beta)$ is not very peaked, as in the case of

[12] A first number of evaluations is required to approximately locate the maximum point y_n^* using a robust Newton's method, but this in not considered here for simplicity.

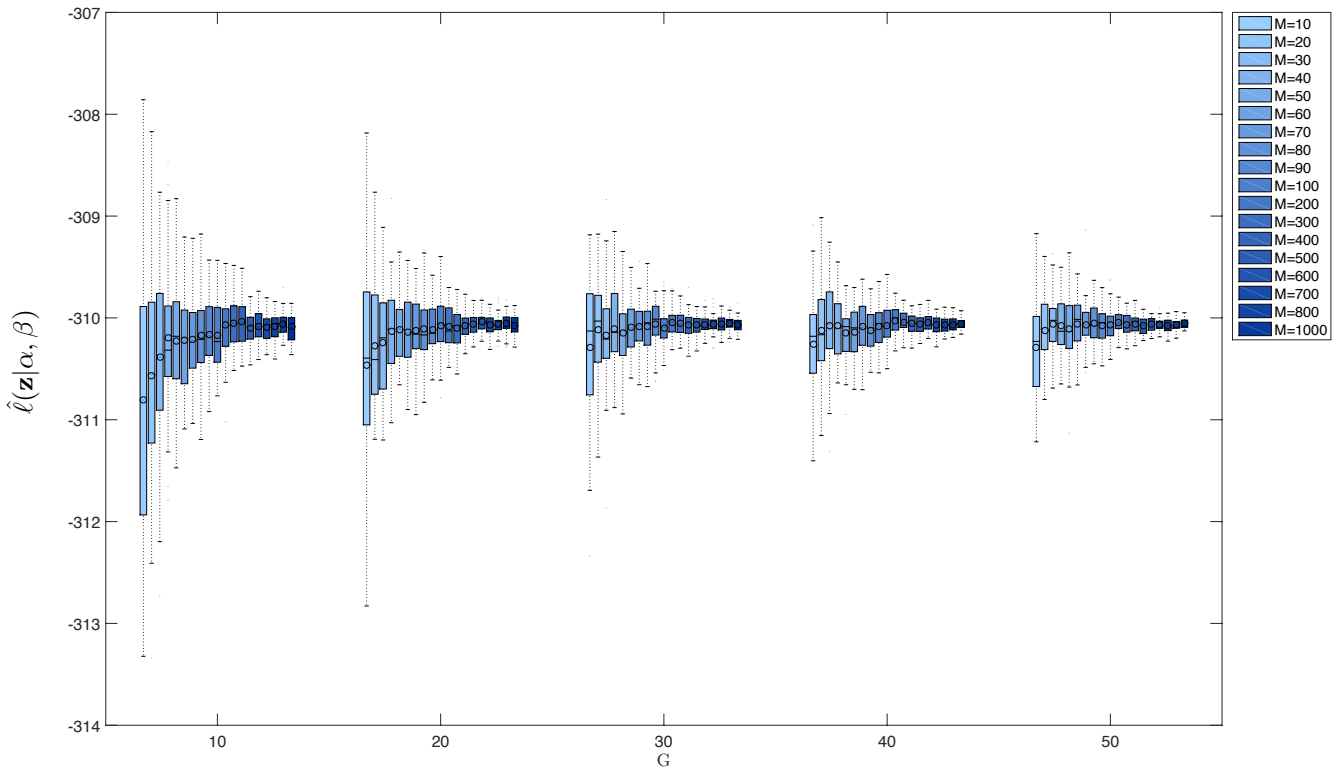


Figure 4.6 Boxplots of logarithms of likelihood estimators $\hat{\ell}(\boldsymbol{z}|\alpha,\beta)$, using the piece-wise constant envelope proposal for the LVs. The $N = 1000$ standard stable data points $\boldsymbol{z}$ are simulated using the CMS algorithm, with $\alpha = 0.5$ and $\beta = 0.7$. For each combination of G and M, the logarithm of the IS based likelihood estimator has been computed 50 times, varying the seed when generating the LVs.

Figure 4.4. Peaked full conditionals are observed when $|z_n|$ is either close to zero, or very large, as in Figure 4.5.[13] As explained in Section 3.2.3, these cases are partially responsible for the high variance of $\hat{p}(\boldsymbol{z}|\alpha,\beta)$, if the parameters G and M are not opportunely tuned. In fact, when both G and M are low and $p(y_n|z_n,\alpha,\beta)$ is peaked, then $R(\boldsymbol{z},\{Y_j\}_{j=1}^M) \approx 0$, where R is the ratio between the likelihood estimate and its value, defined in (3.27). This causes the corresponding PM parameter samplers to have poor mixing (the Markov chains are 'stuck' for many iterations).

A low parameter G can be compensated by sampling a large number of LVs M. Increasing M enables to draw some samples from the high probability region of $p(y_n|z_n,\alpha,\beta)$, thus lowering the variance of the likelihood estimator. However, a very large number of draws might be needed in practice, as shown in Figure 4.6,

[13] In the example in Figure 4.4, the value of z_n is not excessively large, considering that the tail parameter α is small.

that represents the boxplots[14] of the logarithms of the likelihood estimators $\hat{\ell}(\boldsymbol{z}|\alpha,\beta)$ obtained with a piece-wise constant envelope function, varying both G and M. For each combination of G and M, $\hat{\ell}(\boldsymbol{z}|\alpha,\beta)$ has been computed 50 times in a robust way, using (4.16) and (4.17). It is clear that, for small values of G, $\hat{\ell}(\boldsymbol{z}|\alpha,\beta)$ has a large bias (corresponding to large variance of $\hat{p}(\boldsymbol{z}|\alpha,\beta)$), unless the number of LV draws M is larger than 500. A number of similar experiments suggested that a choice of $G = 30$ and $M = 50$ gives a reasonable chain mixing of the PM samplers for any parameter regime.

Generalized Adaptive IS step

Evaluating $p(y_n|z_n,\alpha,\beta)$ might be an expensive operation, in which case it is desirable to minimize L (4.20), the total number of evaluations of this function. In the procedure proposed above, the first G samples are discarded when computing the likelihood estimate $\hat{p}(z_n|\alpha,\beta)$, but they could instead be used for this purpose.

If the cost of adapting the envelope function is small, compared to the cost of evaluating $p(y_n|z_n,\alpha,\beta)$, then it is advisable to keep adapting the proposal distribution for L steps, and use all the non i.i.d. draws, in order to estimate the likelihood. Thus, for each stable data point z_n, $n = 1,\dots,N$, the LVs used in the IS estimate of the likelihood are sampled as follows: Y_1 is uniformly sampled from $q_1(y_n|z_n,\alpha,\beta)$, as in (4.12); the subsequent variables $\{Y_j\}_{j=2}^{L}$ are each sampled from $q_j(y_n|z_n,\alpha,\beta,Y_{1:j-1})$, $j = 1,\dots,L$. Following (3.23a)-(3.23c), it is easy to prove that a likelihood estimator obtained with non i.i.d. samples is positive and unbiased, thus satisfying the PM requirements.

In this case it is also advised to account less for the first samples and more for the last ones, given that the initial proposals distributions (and the relative draws) might be far from the target, while the proposals get closer to the target throughout the adaptations. To obtain this effect, we introduce an increasing sequence $\{r_j\}_{j=1}^{L}$, normalized to sum to 1,

$$r_j = \frac{a_j}{\sum_{k=1}^{L} a_j}, \tag{4.21}$$

with $a_j \geq 0$ and increasing as well; the sequence $\{r_j\}$ substitutes the constant weighting factor $1/M$ in the expression of the likelihood estimator (3.21). We can thus create the following new estimator that accounts for both non i.i.d. samples Y_j

[14] In all the boxplots displayed in this Chapter, the horizontal, the bottom and top edges of the box indicate the 25^{th} and 75^{th} percentiles, respectively; the horizontal line indicates the median, the circle the mean.

and for weighting more the final samples

$$\hat{p}_r(\boldsymbol{z}|\boldsymbol{\theta}) := \prod_{n=1}^{N} \underbrace{\sum_{j=1}^{M} r_j \frac{p(z_n|Y_j,\boldsymbol{\theta})p(Y_j|\boldsymbol{\theta})}{q_j(Y_j|\boldsymbol{\theta},z_n,Y_{1:j-1})}}_{\hat{p}_r(z_n|\boldsymbol{\theta})} = \prod_{n=1}^{N} \hat{p}_r(z_n|\boldsymbol{\theta}), \tag{4.22}$$

where $\boldsymbol{\theta} = (\alpha, \beta)$ and $q_1(y_n|z_n, \alpha, \beta, y_{1:0}) = q_1(y_n|z_n, \alpha, \beta)$, as in (4.12). Observe that (4.22) is easily proved to be unbiased for the true likelihood, thanks to the normalizing condition (3.21). In fact, the adaptive IS estimator in Algorithm 7, is a special case of (4.22) with proposals q_j and weighting sequence

$$r_{1:G-1} = 0, \qquad r_{G:L} = 1/M, \tag{4.23}$$

and the proposal for sampling the last M LVs not being continuously adapted.

We have performed a number of experiments changing the sequence $\{a_j\}$ in (4.21), as shown on the left-hand side of Figure 4.7. The variance of the resulting likelihood estimator (4.22) has been compared to the procedure in Algorithm 7, tuned with a different number of parameters G and M (or, alternatively, G and L). For example, the right-hand side of Figure 4.7 shows a comparison of these strategies for the piece-wise constant envelope proposal, for which it is more likely to see an effect on the variance of the likelihood estimator. Our conclusion is that the linear sequence $a_j = j$ appears to perform marginally better than the other sequences that we examined; however, for our purposes, generalizing the IS does not seem to have a net effect on reducing the variance of the likelihood estimator, if G and M are 'large enough'.

Given the above observation and that, in our case, the (unnormalized) kernel of $p(y_n|z_n, \alpha, \beta)$ is in closed form, and thus not too expensive to evaluate, in the following we choose to focus on the adaptive IS algorithm, rather than using (4.12). However, in Section 4.5.1 we do comparisons with other proposal distributions by fixing the same number L of evaluations of $p(y_n|z_n, \alpha, \beta)$.

4.3.3 Laplace Approximation Proposal

Given the property of unimodality on a compact support of $p(y_n|z_n, \alpha, \beta)$, a choice that appears natural is to use the Laplace approximation of the LV full conditional as proposal distribution $q(y_n|z_n, \alpha, \beta)$ in the IS-PM scheme.

Laplace approximating a target distribution $p(y)$ means fitting a Gaussian log-likelihood to the second order Taylor expansion of the log-target, see e.g. Tierney

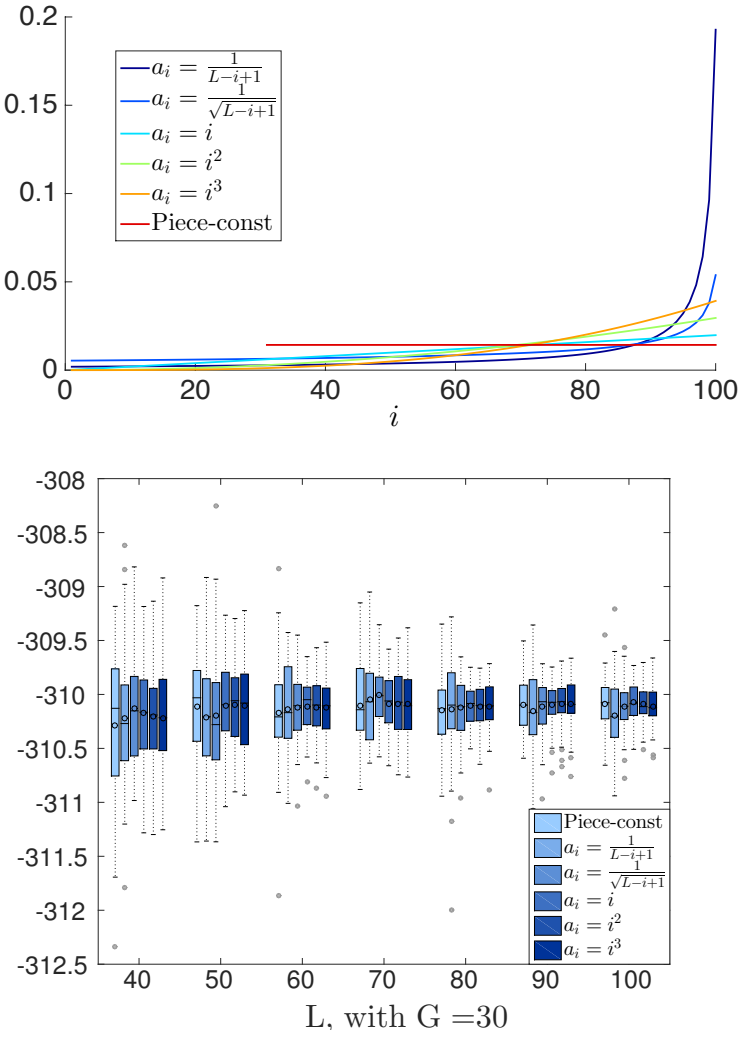


Figure 4.7 Comparison of the standard and generalized adaptive IS based likelihood estimators in the PM scheme. Top: choices of $\{a_i\}$ in the weighting sequence $\{r_i\}$ (4.21); 'Piece-const' is referred to the sequence (4.23). Bottom: Boxplots of logarithms of likelihood estimators (4.22), using the piece-wise constant envelope proposal for the LVs; 'Piece-const' is referred to the estimator obtained through Algorithm 7; the $N = 1000$ standard stable data points $\boldsymbol{z}$ are simulated using the CMS algorithm, with $\alpha = 0.5$ and $\beta = 0.7$.

and Kadane [1986], Robert and Casella [2004] or MacKay [2003]. The approximating PDF is thus a Gaussian whose mode location and curvature match those of the target on a logarithmic scale.

Assuming for simplicity that $p(y)$ is unimodal, with (stationary) maximum point (y^*, p^*), as in the case of $p(y_n|z_n, \alpha, \beta)$, we define $\mathcal{L} := \log p$ to be its logarithm, with first and second derivatives $\mathcal{L}'$, and $\mathcal{L}''$. We also define $u(y)$ to be the Taylor

approximation of $\mathcal{L}$,[15] truncated to the second order, and centred at y^*, as follows

$$u(y) := \mathcal{L}(y^*) + \mathcal{L}'(y^*)(y - y^*) + \frac{1}{2}\mathcal{L}''(y^*)(y - y^*)^2.$$

Given that $\mathcal{L}' = (\log p)' = p'/p$, then y^* is a stationary point also for $\mathcal{L}$, so that

$$u(y) = \mathcal{L}(y^*) + \frac{1}{2}\mathcal{L}''(y^*)(y - y^*)^2,$$

where

$$\mathcal{L}'' = (\log p)'' = \frac{1}{p}\left(-\frac{(p')^2}{p} + p''\right). \tag{4.24}$$

On the other hand, given the PDF of a Gaussian distribution with mean a and variance b^2,

$$n(y|a, b^2) := \frac{1}{\sqrt{2\pi b^2}} \exp\left\{-\frac{1}{2}\frac{(y-a)^2}{b^2}\right\}, \tag{4.25}$$

we can write its logarithm, after re-parametrizing the variance to the precision parameter $\tau := 1/b^2$, as follows

$$\log n(y|a, \tau) = -\frac{1}{2}\log(2\pi) + \frac{1}{2}\log(\tau) + \frac{1}{2}(-\tau)(y - a)^2.$$

Thus, according to the Laplace approximation procedure, we want to determine the parameters a and τ for which the identity $\log n(y|a, \tau) = u(y)$ is satisfied. This leads to

$$\begin{cases} a = y^* & \text{(4.26a)} \\ \tau = -\mathcal{L}''(y^*). & \text{(4.26b)} \end{cases}$$

Observe that, in our case the support of the target distribution, $\mathcal{Y}$ (2.22), is bounded. In order to have a valid IS step in the PM scheme, it is thus necessary to have a proposal with the same support. We thus equate $u(y)$ with the log-PDF of a truncated Gaussian distribution, that has PDF

$$n(y|a, \tau, \mathcal{Y}) = \frac{1}{\Phi(\mathcal{Y})} n(y|a, \tau)\mathbb{1}(y \in \mathcal{Y}), \tag{4.27}$$

[15] Observe that, for the Laplace approximation, it is not relevant if $p(y)$ is normalized or not. In fact, this changes only the constant term in the Taylor expansion of $\mathcal{L}$, but not $\mathcal{L}'$ and $\mathcal{L}''$. The normalizing constant of $p(y)$ needs to be taken into account only if we want to visualize the match on the linear scale.

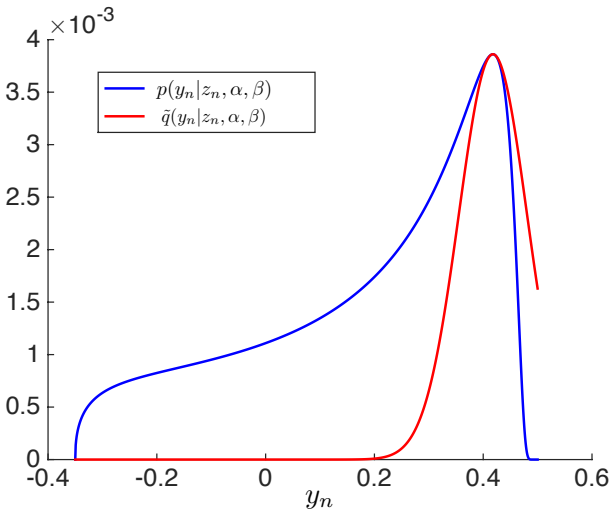


Figure 4.8 Laplace proposal distribution $\tilde{q}(y_n|z_n,\alpha,\beta)$ in the PM sampler, truncated to $\mathcal{Y}_n$, and displayed on a linear scale. The standard stable point is $z_n = 40.83$, with $\alpha = 0.3$ and $\beta = 0.7$.

where n is defined in (4.25), $\mathbb{1}(\cdot)$ is the indicator function, and, with simplified notation, we define $\Phi(\mathcal{Y})$ to be the probability mass assigned by n to $\mathcal{Y}$:

$$\Phi(\mathcal{Y}) := \begin{cases} \Phi(\frac{1/2-a}{b}) - \Phi(\frac{l_{\alpha,\beta}-a}{b}) & z > 0, \\ \Phi(\frac{l_{\alpha,\beta_2}-a}{b}) - \Phi(\frac{-1/2-a}{b}) & z < 0, \end{cases} \tag{4.28}$$

in which $\Phi(\cdot)$ is the CDF of the standard Gaussian distribution. However, truncating the support of the proposal does not affect the validity of the system of equations (4.26a)-(4.26b), because $n(y|a,\tau)$ differs from $n(y|a,\tau,\mathcal{Y})$ only by a normalizing constant (on $\mathcal{Y}$), and this does not affect the mode location and curvature. However, observe that $\Phi(\mathcal{Y})$ needs to be taken into account when computing the logarithm of the likelihood estimator (4.16) obtained with this truncated Gaussian distribution as proposal, and, in fact, it corresponds to the term $Q(z_n,\alpha,\beta)$ in (4.15)

Thus, in order to obtain the Laplace approximation of $p(y_n|z_n,\alpha,\beta)$ (4.5), we need to locate its maximum y_n^* (approximately found using Newton's method to solve (4.9)), and to evaluate the second derivative of its logarithm in y_n^*. Following (4.24), and using (4.10), (4.11) and the stationarity of y_n^*, we have

$$[\log p(y_n|z_n,\alpha,\beta)]''\Big|_{y_n=y_n^*} = -\left[t'(y_n^*)\right]^2,$$

with t' as in (4.6).

Samples from a Gaussian distribution truncated to a domain $\mathcal{Y} = (y_1, y_2)$ can be obtained using the inverse CDF method, see Devroye [1986], Robert [1995]. In particular, if a and b^2 are the mean and the variance of the corresponding not truncated Gaussian, and we define $\bar{y}_1 := (y_1 - a)/b$, and $\bar{y}_2 := (y_2 - a)/b$, and $U \sim \mathcal{U}(0, 1)$ is a uniform RV on the interval $(0, 1)$, then

$$X = \Phi^{-1}\left(\Phi(\bar{y}_1) + U(\Phi(\bar{y}_2) - \Phi(\bar{y}_1))\right) b + a, \tag{4.29}$$

where, as above, Φ is the CDF of the standard normal distribution and Φ^{-1} its inverse. In fact, the term $\Phi(\bar{y}_1) + U(\Phi(\bar{y}_2) - \Phi(\bar{y}_1))$ in the above expression is simply a uniform RV on the interval $(\Phi(\bar{y}_1), \Phi(\bar{y}_2))$.

For example, Figure 4.8 shows the Laplace approximation (in linear scale, once we have taken into account all the normalizing constants) of the same LV full conditional distribution as in Figure 4.4. Observe that the most problematic case now is that of target distributions $p(y_n|z_n, \alpha, \beta)$ 'spread' on $\mathcal{Y}_n$, contrary to the fact that the most difficult case to tackle with adaptive proposals is that of very peaked targets, as in Figure 4.5. This case of spread proposals is thus now responsible for the high variance of $\hat{p}(\boldsymbol{z}|\alpha, \beta)$, because it leads to $R(\boldsymbol{z}, \{Y_j\}_{j=1}^M) \gg 1$, where R is the ratio between the likelihood estimate and its value, defined in (3.27). However, the main difference wrt the adaptive proposals is that, when using the Laplace approximation only M, the number of draws $\{Y_j\}_{j=1}^M$ that determines the quality of the Monte Carlo estimate of the integral of $p(z_n, y_n|\alpha, \beta)$ in the IS step, can be used as tuning parameter. This number might be prohibitively high in order to obtain a good chain mixing, as shown in Section 4.5.1, making the choice of the Laplace approximation not practical for our purposes.

Remark 5. Observe that computing Φ and Φ^{-1} in (4.28) and (4.29) is a numerical operation, that makes our PM scheme with Laplace proposal approximate. However, using truncated Gaussian proposals is a standard procedure in MCMC samplers and the bias introduced by these computations tends to be neglected, because they are done with a high level of precision. In fact, we use truncated Gaussian proposals also for the steps on the parameter presented in the following Section.

4.4 Parameter Priors and Proposals

Once we have defined how to obtain either y_n in the GS or $\hat{p}(z_n|\alpha, \beta)$ in the PM scheme, for each data point z_n, $n = 1, \dots N$, we only need to choose the prior distributions of the parameters and how to draw them at each MCMC iteration, in order to fully specify the two samplers.

We recall that the α-stable CF (2.4) as well as the joint PDF $p_{(Z,Y)(z,y)}$ (2.21) are not continuous for $\alpha = 1$, and some of the analytical expressions used so far are not well defined for $\alpha = 0$ (a value generally not valid when dealing with stable laws). For this reason, in the GS Buckle [1995] adds a constraint on the prior and the proposal distributions, requiring that $\alpha \in S_\alpha$, where, in turn, $S_\alpha \in \{(0.1, 0.9), (1.1, 2)\}$. We follow this choice in our PM sampler, by further bounding the support of the prior and proposal distribution of β to $S_\beta \in \{(-1, 0), (0, 1)\}$, for simplicity.[16] Specifically, in our simulations we make sure that S_α and S_β contain (α_T, β_T), the values of the parameters used to generate the stable points dataset. In real scenarios, where (α_T, β_T) are not known, the choice of S_α and S_β could be guided by inspection of the dataset and based on the meaning of the parameters (we might know whether or not the mean of the dataset is likely to exist, and in which direction the data are likely to be skewed).

For simplicity, we assume that the parameters are *a priori* independent and uniformly distributed on $S_\alpha \times S_\beta$, with joint prior PDF

$$p(\alpha, \beta) = \frac{1}{|S_\alpha|}\mathbb{1}(\alpha \in S_\alpha) \times \frac{1}{|S_\beta|}\mathbb{1}(\beta \in S_\beta) = p(\alpha)p(\beta), \tag{4.30}$$

where $|S|$ denotes the Lebesgue measure of the set $S \in \mathbb{R}$, and $\mathbb{1}(\cdot)$ is the indicator function. We refer to Buckle [1995] for remarks on different possible choices of priors (e.g. Beta distributions instead of uniforms) and on the dependence between the parameters (e.g. when α tends toward 2, β has less influence). We also recall that Buckle [1995] deals also with inference of the parameters σ and μ in non-standardized datasets, in which case a uniform prior on μ is not proper, but does lead to proper posteriors, while a conjugate prior is chosen for σ.

As anticipated in Section 3.3.1, in the GS the full-conditional distributions of the parameters are not sampled from directly, but targeted with a MH-GS step. According to Buckle [1995], a re-parametrization of the LV (from y_n to $v_n = t_{\alpha,\beta}(y_n)$, with $t_{\alpha,\beta}$ defined in (2.8)) seems to be necessary to enable the mixing of Markov chains. Our simulations, reported in Riabiz et al. [2015], confirm this fact: we were not able to produce a converging MH-GS scheme that targets the original full-conditionals. Thus, following (3.20), and using (4.30), we can write the target

[16] However, we believe that these constraints could be relaxed, by choosing a sampler able to 'jump' across the domain boundaries. We expect that such a sampler would be effective especially in the PM scheme, that does not suffer from the correlation between the LVs and the parameters.

distribution of the k^{th} MH step for α as

$$p(\alpha^{(k)}|\beta^{(k-1)}, \boldsymbol{v}^{(k-1)}, \boldsymbol{z}) \propto p(\boldsymbol{z}, \boldsymbol{v}^{(k-1)}|\alpha^{(k)}, \beta^{(k-1)})p(\alpha^{(k)}),$$

and a similar expression yields for the target of the MH step for β. This re-parametrization both adds computational complexity, and it introduces bias in the sampler. In fact, due to the change of variables from $\boldsymbol{y}$ to $\boldsymbol{v}$, in order to evaluate the MH acceptance probability, we need to compute the Jacobian of the inverse transformation $t^{-1}_{\alpha,\beta}$,

$$J = \left| \frac{\mathrm{d}t_{\alpha,\beta}}{\mathrm{d}y} \right|^{-1}_{y=t^{-1}_{\alpha,\beta}(v)},$$

where we remark that the set $\{y_n\}_{n=1}^N = \{t^{-1}_{\alpha,\beta}(v_n)\}_{n=1}^N$ is different from $\left\{y_n^{(k)}\right\}_{n=1}^N$ each time a new value of the parameters α or β is proposed. As explained in Section 4.1, the inversion of $t_{\alpha,\beta}$ is performed numerically, for example with a safeguarded Newton scheme. We refer to Buckle [1995] for considerations on how to speed the numerical procedure; however, the fact that a numerical scheme is used implies that the MH acceptance probabilities are not exactly evaluated, making the GS further biased.

Even when using the new LVs $\boldsymbol{v} := [v_1, \ldots, v_N]$ in the parameter full-condtional distributions, our implementation of the GS scheme differs from that presented by Buckle [1995]. In the literature, the problem of the selection of the parameter proposal distributions is alleviated by building adaptive rejection sampling envelopes, according to the methodology suggested by Gilks and Wild [1992] and its extensions for non log-concave distributions. However, this adds unnecessary computational complexity to the sampling procedure,[17] so we prefer using simpler proposal distributions in this evaluation. In particular, we adopt proposal distributions for the parameters that are a tunable mixture between a truncated Gaussian random walk (with probability p) and an independent move (with probability $1-p$). For example, the proposal for α is

$$q_\alpha(\alpha|\alpha^{(k)}) = \begin{cases} n(\alpha|\alpha^{(k)}, 1/\sigma_\alpha^2, S_\alpha) & \text{with probability } p_\alpha, \\ \dfrac{1}{|S_\alpha|}\mathbb{1}(\alpha \in S_\alpha) & \text{with probability } 1-p_\alpha, \end{cases} \tag{4.31}$$

where $n(y|a, \tau, \mathcal{Y})$ is the PDF of a truncated Gaussian distribution, defined in (4.27), and to which (4.31) assigns mean $\alpha^{(k)}$ and precision parameter $1/\sigma_\alpha^2$. The first

[17] Observe, however, that this shows the point that it would be possible to circumvent the bias introduced by using truncated Gaussian proposals for the parameters.

Algorithm 8 PM-IS based MH within Gibbs for (α, β), given $\boldsymbol{z} \sim \mathcal{S}_\alpha(1, \beta, 0)$

1: Set $(\alpha^{(0)}, \beta^{(0)})$ arbitrarily

2: Compute $\zeta^{(0)} \leftarrow \mathrm{IS}(\alpha^{(0)}, \beta^{(0)}, \boldsymbol{z})$

3: **for** $k = 1, \ldots, N_{it}$ **do**

1. Given $(\alpha^{(k-1)}, \beta^{(k-1)}, \zeta^{(k-1)})$
   (a) Propose $\alpha' \sim q_\alpha(\cdot \,|\, \alpha^{(k-1)})$
   (b) Compute $\zeta' \leftarrow \mathrm{IS}(\alpha', \beta^{(k-1)}, \boldsymbol{z})$
   (c) Set

$$\left\{\alpha^{(k)}, \zeta''\right\} := \begin{cases} \{\alpha', \zeta'\} & \text{with prob. } \rho_\alpha \\ \left\{\alpha^{(k-1)}, \zeta^{(k-1)}\right\} & \text{else,} \end{cases}$$

   where $\rho_\alpha =$

$$1 \wedge \frac{\zeta'}{\zeta^{(k-1)}} \frac{p(\alpha', \beta^{(k-1)})}{p(\alpha^{(k-1)}, \beta^{(k-1)})} \frac{q_\alpha(\alpha^{(k-1)} \,|\, \alpha')}{q_\alpha(\alpha' \,|\, \alpha^{(k-1)})}$$

2. Given $(\alpha^{(k)}, \beta^{(k-1)}, \zeta'')$
   (a) Propose $\beta' \sim q_\beta(\cdot \,|\, \beta^{(k-1)})$
   (b) Compute $\zeta''' \leftarrow \mathrm{IS}(\alpha^{(k)}, \beta', \boldsymbol{z})$
   (c) Set

$$\left\{\beta^{(k)}, \zeta^{(k)}\right\} := \begin{cases} \{\beta', \zeta'''\} & \text{with prob. } \rho_\beta \\ \left\{\beta^{(k-1)}, \zeta''\right\} & \text{else,} \end{cases}$$

   where $\rho_\beta =$

$$1 \wedge \frac{\zeta'''}{\zeta''} \frac{p(\alpha^{(k)}, \beta')}{p(\alpha^{(k)}, \beta^{(k-1)})} \frac{q_\beta(\beta^{(k-1)} \,|\, \beta')}{q_\beta(\beta' \,|\, \beta^{(k-1)})}$$

component corresponds to a local exploration of the state space, and requires the tuning of the variance σ_α^2, while the second one represents the attempt at a more global move. An analogous proposal $q_\beta(\beta|\beta^{(k)})$ can be used for the skeweness parameter, with hyper-parameters σ_β^2 and p_β.

Remarkably, in the PM scheme the problem of re-parametrizing the LVs for sampling the parameters does not subsist, because the likelihood estimator ζ takes the role that the true likelihood has in a simple MH or GS, and we can thus sample from or target the parameter full-conditionals conditioned on the original LVs $\boldsymbol{y}$. In particular, the PM Algorithm 6 could be used either directly, with $\boldsymbol{\theta} = (\alpha, \beta)$, or as a

Metropolis-within-Gibbs step, as detailed in Algorithm 8. We refer to the first scheme as PM-MH and to the second one as PM-MH-GS. In our case, the PM-MH-GS has twice the simulation time of the PM-MH sampler, since ζ is computed twice per each iteration (once to evaluate the acceptance probability for α, and once to evaluate that for β). However, tuning of PM-MH-GS is simpler than for PM-MH, because the two parameters α and β are proposed and accepted or rejected separately, implying that proposal distributions of the form (4.31) can be used, while a bivariate proposal needs to be chosen in the PM-MH scheme. Thus we choose to implement the PM-MH-GS sampler, as shown in the following section.

4.5 Simulation Results and Comparisons

We conclude this chapter by reporting numerical simulations aimed at comparing the choice of proposal distributions in the PM sampler, as well as the results obtained by running a GS and PM scheme on the same stable dataset.

4.5.1 Comparison of LV Proposals

In order to compare the proposal distributions for the LVs presented above, we examine the variance of the relative likelihood estimator, namely the bias of its logarithm $\hat{\ell}(\boldsymbol{z}|\alpha,\beta)$ (4.16). In particular, we compare the following proposals, by using the same number L of evaluations of each of the LVs full conditional distribution $p(y_n|z_n,\alpha,\beta)$:

- uniform proposal: this corresponds to the piece-wise constant envelope proposal from Section 4.3.1, without adaptation ($G = 1$). In this case L is also the number of LVs drawn to form each log-likelihood summand $\hat{\ell}(z_n|\alpha,\beta)$;
- piece-wise constant envelope proposal as in Section 4.3.1, with $G = 30$ adaptations. Thus $L - G$ is the number of i.i.d. LVs drawn to form $\hat{\ell}(z_n|\alpha,\beta)$;
- piece-wise linear interpolating proposal as in Section 4.3.1, with $G = 30$ adaptations. Thus $\hat{\ell}(z_n|\alpha,\beta)$ is obtained with $L - G$ i.i.d. draws;
- Laplace approximation proposal, truncated on $\mathcal{Y}_n$, as in Section 4.3.3, and from which L i.i.d. LVs are drawn to obtain $\hat{\ell}(z_n|\alpha,\beta)$.

Figure 4.9 shows the boxplots of the logarithms of the likelihood estimators obtained with the above proposals, $N = 100$ standardized stable points with $\alpha = 0.5$ and $\beta = 0.7$. The boxplots are relative to 50 realizations of the likelihood estimator.

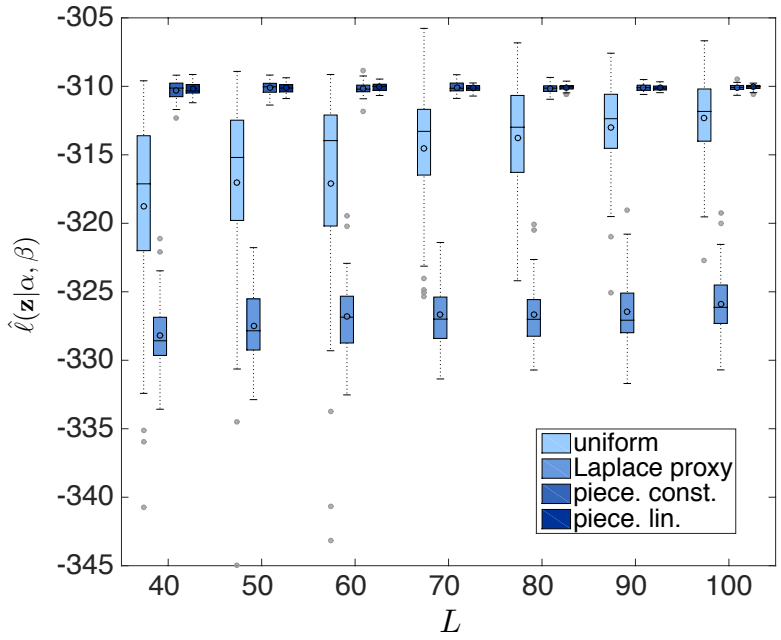


Figure 4.9 Boxplots of logarithms of likelihood estimators $\hat{\ell}(\boldsymbol{z}|\alpha,\beta)$ using four different proposals for the LVs in the IS step (see the text for the details). The same number L of evaluations of the LVs full conditional distribution $p(y_n|z_n,\alpha,\beta)$ is used in the comparison. The $N = 100$ standard stable data points $\boldsymbol{z}$ are simulated using the CMS algorithm, with $\alpha = 0.5$ and $\beta = 0.7$.

The results obtained are representative of the situation faced for a vast range of parameters examined. In particular, it is possible to observe that the estimators $\hat{\ell}(\boldsymbol{z}|\alpha,\beta)$ resulting from the piece-wise constant and linear adaptive proposal have smaller bias (higher mean) than the other estimators; the estimator obtained from the Laplace approximation performs significantly poorly. Furthermore, we underscore that increasing the order of the piece-wise IS proposal distribution decreases the bias of $\hat{\ell}(\boldsymbol{z}|\alpha,\beta)$, especially for low values of G.

We attribute the poor performance of the estimators obtained with uniform and Laplace approximation proposals to the fact that situations similar to those in Figures 4.5 and 4.8 arise, respectively. In particular, in the first case it happens that $q(Y_j|z_n,\alpha,\beta) \gg p(Y_j|z_n,\alpha,\beta)$, while in the second $q(Y_j|z_n,\alpha,\beta) \ll p(Y_j|z_n,\alpha,\beta)$, for some of the draws $\{Y_j\}_{j=1}^{M}$. In both cases, increasing the number of samples M decreases the bias of $\hat{\ell}(\boldsymbol{z}|\alpha,\beta)$, as shown in Figure 4.10, that is related to the same standard stable dataset of Figure 4.9 ($N = 100$, $\alpha = 0.5$, $\beta = 0.7$). However, in both cases, a very large number of samples (prohibitively large with Laplace approximation proposal) has to be drawn to achieve results similar to those with adaptive proposals.

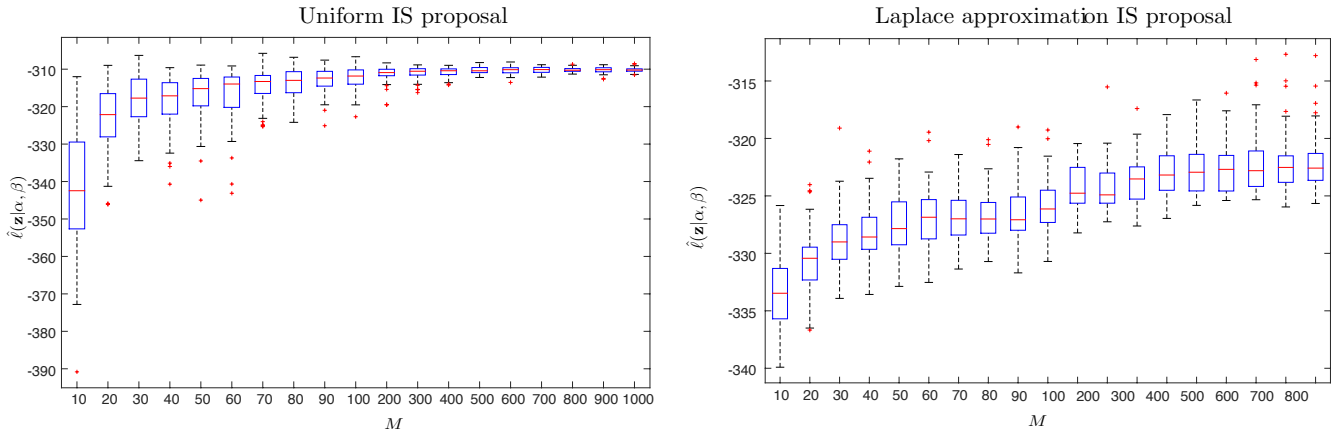


Figure 4.10 Boxplots of logarithms of likelihood estimators with Uniform and Laplace approximation proposals for the LVs in the IS step, increasing M (namely L). The $N = 100$ standard stable data points $\boldsymbol{z}$ are simulated using the CMS algorithm, with $\alpha = 0.5$ and $\beta = 0.7$.

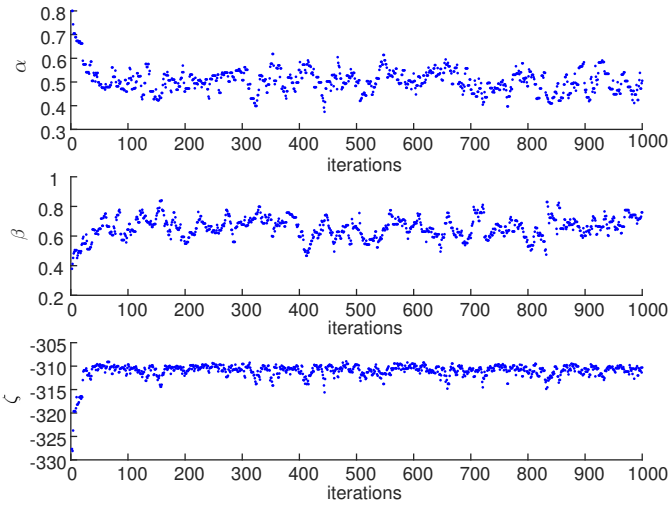
This is reflected in the mixing of the PM-MH-GS Markov chains, related to the same stable dataset, and obtained with the four different proposals, as shown in Figure 4.11. Good chain mixing is obtained with the adaptive proposals, while a large number of draws ($M = 1000$) is required for the PM sampler with uniform proposals to mix well. Finally, not even $M = 1000$ draws are sufficient for the PM scheme with Laplace approximation proposals, in which the Markov chains are 'stuck' on the same parameters values for many iterations, and, most remarkably, they have not converged to the 'true' values of the parameters, used to generate the dataset.

4.5.2 Comparison of the PM and the GS Schemes

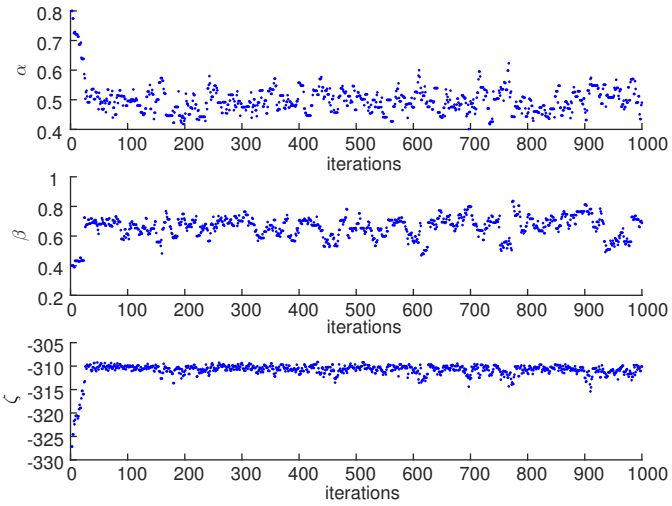
In this section we compare the performance of the GS scheme by Buckle [1995], and of our proposed PM-MH-GS scheme from Algorithm 8, that samples separately the parameters α and β. Given the above considerations, we decide to use the piece-wise linear proposal, with $G = 30$ and $M = 50$ (we increase G here to have a better likelihood estimator).

We perform simulations for a standard stable dataset, with now $N = 300$ points, and with $\alpha = 0.5$ and $\beta = 0.7$, corresponding to a slow decay of the tails, and a positively skewed distribution. The respective initial values of the chains are $\alpha^{(0)} = 0.8$ and $\beta^{(0)} = 0.4$, and the GS and PM-MH-GS algorithms are run for $N_{it} = 5000$ iterations. Moreover, we use $p_\alpha = p_\beta = 0.85$, and $\sigma^2_\alpha = \sigma^2_\beta = 10^{-3}$ for the parameters proposals (4.31).

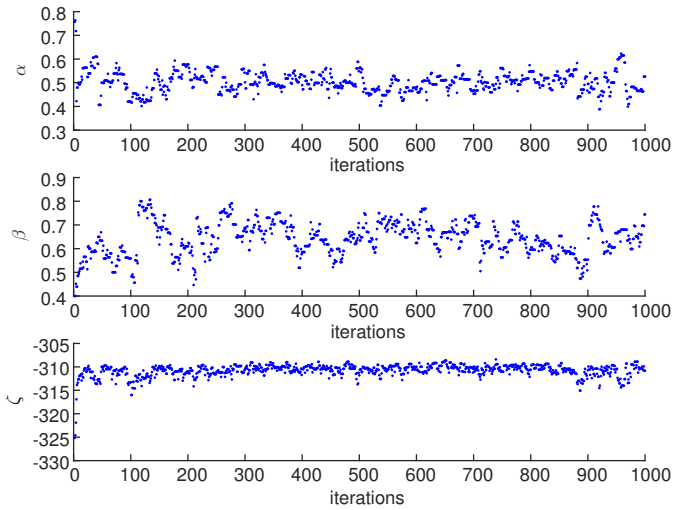
In Figure 4.12a we show the trace plots of the chains for α and β. We remark that the conditional GS suffers from correlation between the parameters and the LVs $\boldsymbol{y}$,



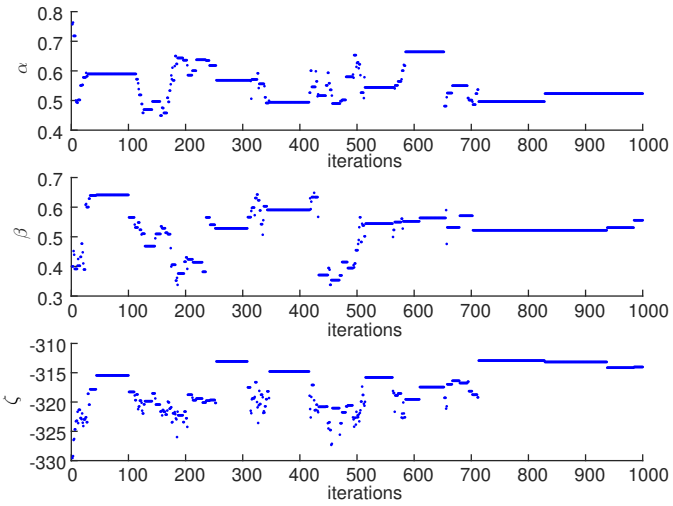
(a) Piece-wise constant ($G = 20, M = 50$)



(b) Piece-wise linear ($G = 20, M = 50$)

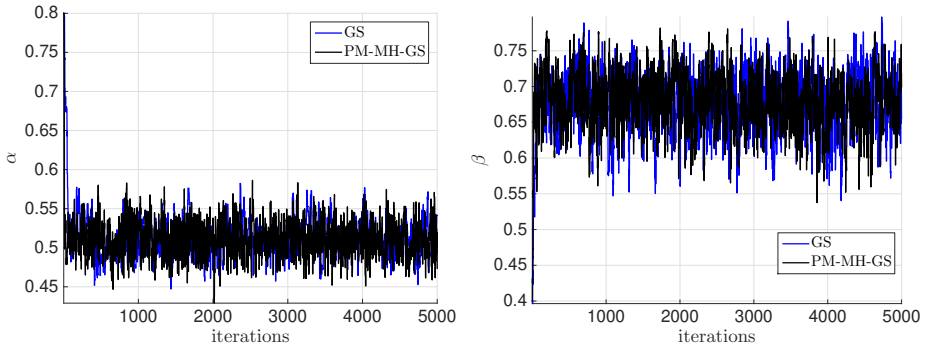


(c) Uniform ($M = 1000$)

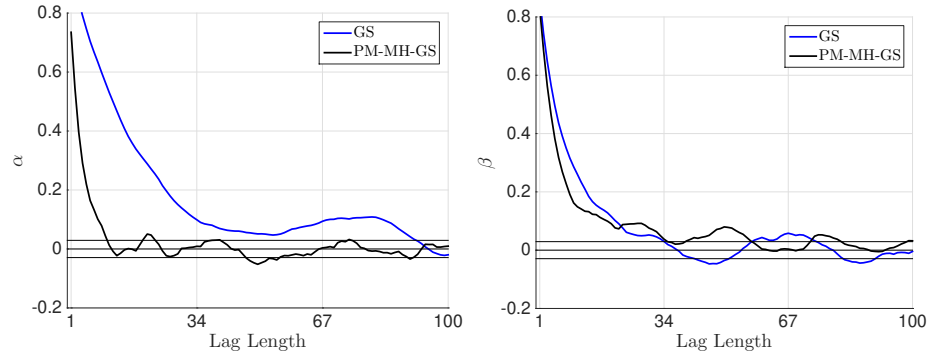


(d) Laplace approximation ($M = 1000$)

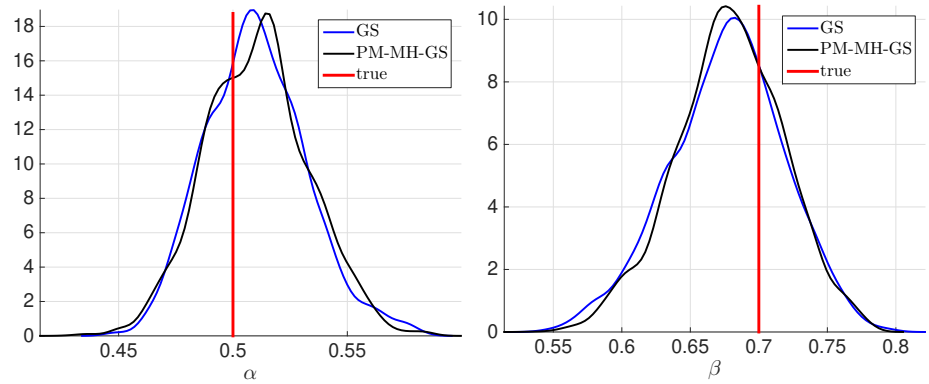
Figure 4.11 Markov chains on the parameters α and β and the likelihood $\zeta = \hat{p}(\boldsymbol{z}|\alpha, \beta)$, in the PM sampler using four different proposals for the LVs, specified in the subcaptions. The $N = 100$ standard stable data points $\boldsymbol{z}$ are simulated using the CMS algorithm, with $\alpha = 0.5$ and $\beta = 0.7$. The initial values for the parameters are $\alpha^{(0)} = 0.5$, $\beta^{(0)} = 0.4$. The proposals for the parameters are more local than global, with probabilities of truncated Gaussian random walk moves $p_\alpha = p_\beta = 0.85$, and $\sigma_\alpha^2 = \sigma_\beta^2 = 10^{-3}$ in (4.31).



(a) Trace plots



(b) Autocorrelation functions



(c) Posterior estimates

Figure 4.12 Trace plots (top), sample autocorrelation functions (centre) and empirical posterior distributions (bottom) of α (left) and β (right), varying the sampling scheme (GS versus PM-MH-GS). See the text for the setting of the hyperparameters.

making it difficult to find a tuning of the proposals that leads to convergent Markov chains. This effect is successfully reduced by means of the re-parametrization to the LVs $\boldsymbol{v}$, as explained in Section 4.4, and displayed in the Figure. The PM method with the chosen G and M achieves, on the other hand, similar performances to the GS scheme, without the necessity of transforming the latent variables. Because of this, we can also conclude that the bias introduced by the re-parametrization of the

LVs in the GS, a formal drawback of this scheme, seems to have a negligible effect, at least for the set of parameters that we examined.

The sample autocorrelation functions (3.15) are displayed in Figure 4.12b, and they appear comparable, meaning that the two schemes generate Markov chains with a similar effective sample size. An analysis based on the lag corresponding to the first crossing of the rejection band (black horizontal lines), indicates that decreasing G, the number of adaptations of the LV proposal, or M, the number of draws from it, decreases the effective sample size of the PM scheme.

Finally, in Figure 4.12c we show kernel density estimates of the empirical posterior distributions (3.14) of the parameters, produced with the two sampling schemes, after a 500 iterations burn-in. The red vertical lines represent the true values used to generate the sample; the posterior means obtained with both samplers are $\mathbb{E}[\alpha|\boldsymbol{z}] = 0.51$, $\mathbb{E}[\beta|\boldsymbol{z}] = 0.68$, revealing a small difference in precision between the two methods.

Chapter 5

A Central Limit Theorem for the PSR Residual

In this and the following two Chapters, we report results on the MaSMiN model for stable RVs deriving from the PSR, that was introduced in Section 2.4.3.

As shown in Section 3.3.4, the PSR and the MaSMiN representation of the stable distribution are extremely relevant for inference purposes, because they enable to apply to the stable law methods that are valid for conditionally Gaussian distributions. However, due to the series structure of this representation, it is not possible to use the exact MaSMiN model, because it involves infinite sequences of LVs. Given that mere truncations of the series 'converge slowly',[1] in the recent literature two possible approximations of the residuals of the series have been presented, the GAA and the GAMA. As recalled in Section 2.4.4, both approximations assume Gaussianity of the residual of the series that they consider. Nonetheless, to the best of our knowledge this assumption has neither been formally proven nor the amount of approximation error has been quantified.

Thus, the main contribution of the current Chapter is to present and prove a CLT for the PSR residual in the GAA scheme, a result that holds asymptotically, as the truncation limit of the series is set to infinity. The structure of the Chapter is as follows: first the moments of the residuals are characterized, using an appropriate limiting technique; then the Taylor series expansion of the CF of the standardized residual is shown to converge to that of the standard Gaussian, and this is equivalent to the statement of the CLT. Furthermore, we obtain expressions for the CF of the residual and the PSR, in the case the RVs $\{W_j\}$ are Gaussian, and this is the second contribution of this Chapter. These expressions will be used in Chapters 6 and 7 to

[1] The convergence metrics (or, alternatively, divergence measure) that we consider are specified in Section 5.3, while results devoted to the study of the approximation of the PSR with its truncation are recalled in Section 7.3.

formulate bounds on the error that is introduced by the non-asymptotic application of the Gaussian approximation of the residual, that is necessarily made in practice.

Throughout this and the following two Chapters, the sequence $\{\Gamma_j\}$ will denote the successive arrival times of a unit rate Poisson process. If a RV X is defined as a series of random terms involving the sequence $\{\Gamma_j\}$, then $X_{(c,d)}$ will denote the sum of those terms corresponding to indices j such that $\Gamma_j \in (c,d)$, for $0 \leq c < d \leq \infty$. The number of terms in $X_{(c,d)}$ is denoted by $N_{(c,d)}$, with the convention that $X_{(c,d)} = 0$ if $N_{(c,d)} = 0$. A subscript notation is used for the moments of such RVs, e.g. $m_{(c,d)}$ is the mean of $X_{(c,d)}$.

Recall that the PSR induces a re-parametrization between the stable law parameters β and σ and the moments of the i.i.d. variables $\{W_j\}$ (2.31), through (2.34b) and (2.34a). In order to obtain the MaSMiN LV model for the stable distribution (2.38a)-(2.38b), we are interested in the case $W_1 \sim \mathcal{N}(\mu_W, \sigma_W^2)$, which induces a mapping between (μ_W, σ_W^2) and (σ, β). However, our proof of the CLT for the PSR residual holds for the class of distributions on W_1 satisfying $\mathbb{E}[|W_1|^3] < \infty$,[2] and that encompasses any Gaussian distribution for W_1. Thus, in the first part of this Chapter we relax the assumption $W_1 \sim \mathcal{N}(\mu_W, \sigma_W^2)$. This assumption is retrieved in the second part of this Chapter, and in the following Chapters, where we further assume that $\beta = 0$ in order to prove our results.

Finally, we will work with $\alpha \neq 1$, because the expressions of the quantities involved are different for $\alpha = 1$.

Part of the results *have been developed in collaboration with Dr. Tohid Ardeshiri and Prof. Ioannis Kontoyiannis*, and they were presented in Riabiz et al. [2017a,b, 2018a].

5.1 Truncation of the PSR for the GAA Scheme

We start by setting the notation that is used in the GAA for the PSR residual, where we recall that the PSR is the random series representation of a stable RV $X \sim \mathcal{S}_\alpha(\sigma, \beta, 0)$, given in (2.29). As summarized in Section 2.4.4, the GAA is an approximation scheme, firstly introduced by Lemke [2014] and related literature, that aims at probabilistically accounting for the PSR residual. In particular, given that the PSR is stochastically decaying, the series is truncated to a part that includes the first summands corresponding to the indices j s.t. $\Gamma_j < c$, where the sequence

[2] By Hölder's inequality, the condition $\mathbb{E}[|W_1|^3] < \infty$ implies $\mathbb{E}[|W_1|^\alpha] < \infty$ for $0 < \alpha < 2$, and hence it further guarantees that the variables $\{W_j\}$ satisfy the PSR requirement (2.31).

$\{\Gamma_j\}$ is given in (2.30), and $c > 0$ is a truncation constant. The distribution of the residual term of the series is approximated by an 'appropriately chosen' Gaussian RV, an approximation that we formally justify in Section 5.2.

Thus, as illustrated at the top of Figure 5.1, we split the PSR into

$$X = X_{(0,c)} + R_{(c,\infty)}, \tag{5.1}$$

where $X_{(0,c)}$ is the truncated PSR

$$X_{(0,c)} := \sum_{j:\Gamma_j \in [0,c]} W_j \Gamma_j^{-1/\alpha}, \tag{5.2}$$

and $R_{(c,\infty)}$ is the PSR residual term, defined as

$$R_{(c,\infty)} := \lim_{d\to\infty} R_{(c,d)}, \quad d > c, \tag{5.3}$$

with

$$R_{(c,d)} := \sum_{j:\Gamma_j \in (c,d)} W_j \Gamma_j^{-1/\alpha} - \mathbb{E}[W_1] \sum_{j=1}^{\lfloor d \rfloor} b_j^{(\alpha)}, \tag{5.4}$$

and $\lfloor \cdot \rfloor$ denoting the lower integer part. To simplify the notation, from now on we will assume that d is integer so that $\lfloor d \rfloor = d$. Notice that the limit in (5.3) denotes convergence in distribution, and it is guaranteed to exist by the fact that the full series converges to an α-stable RV, according to the PSR (2.29), and by the fact that $X_{(0,c)}$ is a well defined RV, because it is the sum of an almost surely finite number of random summands.

In order to state and prove the CLT in Section 5.2, we need to determine the constants that are used to centre and normalize the residual $R_{(c,\infty)}$ (5.3). We find these constants by deriving the exact first and second order moments of $R_{(c,d)}$ (5.4) and then take their limit, as $d \to \infty$. Observe that the resulting expressions are not guaranteed to be the moments of the true residual $R_{(c,\infty)}$, because we do not prove that the conditions necessary to exchange limits with integrals (expectations) hold. However, the values that we obtain with this approach are 'appropriate': in the proof of the CLT, the residual, accordingly standardized, does converge to the standard Gaussian distribution.

We proceed as follows. According to the properties of Poisson processes, see Kingman [1992], conditioned on the Poisson number of events, the ordered arrival times $\{\Gamma_j\}$ may equivalently be written as an unordered set of i.i.d. *uniformly* distributed random variables $\{U_j\}$. This implies that we can write the following

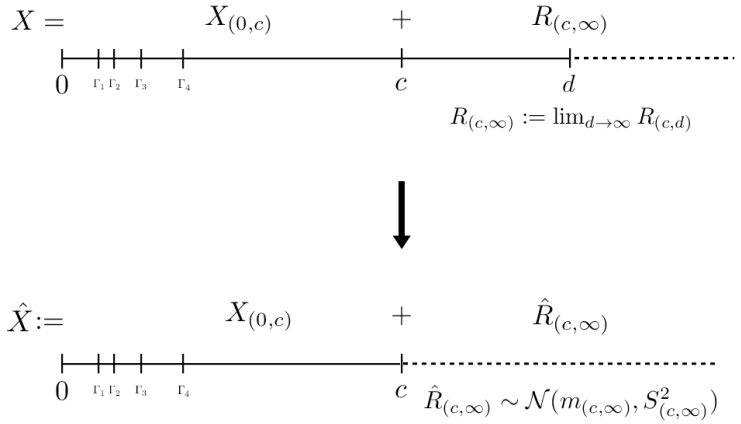


Figure 5.1 Top: illustration of the Poisson series representation of the stable RV X, and its truncation in terms of the set $\{\Gamma_j \leq c\}$; $X_{(0,c)}$ and $R_{(c,d)}$ are defined in (5.2) and (5.4), respectively. Bottom: $\hat{X}$ is our proposed approximation of X, where, according to Theorem 5.2.1, we approximate the PSR residual with the Gaussian RV $\hat{R}_{(c,\infty)}$ with parameters (5.17) and (5.18).

generative model for $R_{(c,d)}$:

$$R_{(c,d)} = \sum_{j=1}^{N_{(c,d)}} Y_j - B, \tag{5.5}$$

$$N_{(c,d)} \sim \text{Poisson}(d-c),$$

$$Y_j := W_j U_j^{-1/\alpha}, \tag{5.6}$$

$$W_j \sim F,$$

$$U_j \sim \mathcal{U}(c,d), \tag{5.7}$$

$$B := \mathbb{E}[W_1]\frac{\alpha}{\alpha-1} d^{\frac{\alpha-1}{\alpha}} \mathbb{1}(\alpha \in (1,2)), \tag{5.8}$$

where Poisson$(d-c)$ is the Poisson distribution with mean $d-c$, F is a distribution on W_1 which is assumed to satisfy (2.31), and $\mathcal{U}(c,d)$ is the uniform distribution on (c,d). The expression for B is obtained from (2.33). In summary, we can think about $R_{(c,d)}$ as a compound Poisson process, containing two sources of randomness: the random Poisson number of arrivals, $N_{(c,d)}$, and the random distribution of the variables being summed, Y_j. Based on this observation it is possible to prove the following lemma.

Lemma 2. Let Y_1 be defined as in (5.6), with $\mathbb{E}[W_1^2] < \infty$,[3] and let $\phi_{Y_1}(s)$ be its CF. Let also B be defined as in (5.8). Then $\phi_{R_{(c,d)}}(s)$, the CF of $R_{(c,d)}$, is

$$\phi_{R_{(c,d)}}(s) = \exp\Big((d-c)\big(\phi_{Y_1}(s)-1\big) - isB\Big). \tag{5.9}$$

Moreover, $m_{(c,d)}$, and $S^2_{(c,d)}$, the mean and variance of the residual $R_{(c,d)}$, are given by:

$$m_{(c,d)} = \mathbb{E}[W_1]\frac{\alpha}{\alpha-1}\left(d^{\frac{\alpha-1}{\alpha}} - c^{\frac{\alpha-1}{\alpha}}\right) - B, \tag{5.10}$$

$$S^2_{(c,d)} = \mathbb{E}[W_1^2]\frac{\alpha}{\alpha-2}\left(d^{\frac{\alpha-2}{\alpha}} - c^{\frac{\alpha-2}{\alpha}}\right). \tag{5.11}$$

Proof: We make use of the observation that $R_{(c,d)}$ can be viewed as a compound Poisson process. Hence, to compute expectations with respect to its distribution, we apply the 'law of total expectation': we first condition on $N_{(c,d)}$, the random number of terms in (c,d), and we then marginalize out $N_{(c,d)}$. Then, using the model (5.5) for $R_{(c,d)}$

$$\begin{aligned}
\phi_{R_{(c,d)}}(s) &= \mathbb{E}\Big[\exp\big(isR_{(c,d)}\big)\Big] \\
&= \mathbb{E}\Big[\mathbb{E}\Big[\exp\Big(is\Big(\sum_{j=1}^{N_{(c,d)}} W_j\Gamma_j^{-1/\alpha} - \mathbb{E}[W_1]\sum_{j=1}^{d} b_j^{(\alpha)}\Big)\Big)|N_{(c,d)}\Big]\Big] \\
&= \mathbb{E}\Big[\mathbb{E}\Big[\exp\Big(is\Big(\sum_{j=1}^{N_{(c,d)}} W_jU_j^{-1/\alpha} - \mathbb{E}[W_1]\sum_{j=1}^{d} b_j^{(\alpha)}\Big)\Big)|N_{(c,d)}\Big]\Big] \\
&= \mathbb{E}\left[\mathbb{E}\left[e^{is\sum_{j=1}^{N_{(c,d)}} Y_j}\middle| N_{(c,d)}\right]\right] \times \exp\Big(-is\mathbb{E}[W_1]\sum_{j=1}^{d} b_j^{(\alpha)}\Big) \\
&= \sum_{n=0}^{\infty}\mathbb{E}\left[e^{is\sum_{j=1}^{n} Y_j}\right]\mathbb{P}\left(N_{(c,d)}=n\right) \times \exp\left(-isB\right) \\
&= \sum_{n=0}^{\infty}(\phi_{Y_1}(s))^n\frac{(d-c)^n}{n!}e^{-(d-c)} \times \exp\left(-isB\right) \\
&= \exp\left((d-c)(\phi_{Y_1}(s)-1) - isB\right),
\end{aligned}$$

where $\phi_{Y_1}(s)$ is the CF of $Y_1 = W_1U_1^{-1/\alpha}$.

If the moments of a given RV V are finite, then they are related to the derivatives in 0 of its CF as follows

$$\mathbb{E}\left[V^k\right] = (-i)^k\phi_V^{(k)}(0), \tag{5.12}$$

[3] This is again not in conflict with the PSR requirement $\mathbb{E}[|W_1|^\alpha] < \infty$, thanks to Hölder's inequality.

see e.g. [Feller, 1966, Lemma 2, p. 512]. Now, since we assume $\mathbb{E}[W_1^2] < \infty$, then $\mathbb{E}[Y_1^2]$ is finite, and hence $R_{(c,d)}$ has finite first and second moments, given that it contains the sum of an almost surely finite number of variables Y_j. Under this assumption, from (5.12) and the above expression of $\phi_{R_{(c,d)}}(s)$ we readily obtain the mean and the variance of $R_{(c,d)}$ as a function of the moments of Y_1 as

$$\begin{aligned} m_{(c,d)} &= \mathbb{E}[R_{(c,d)}] \\ &= (-i)\phi'_{R_{(c,d)}}(0) \\ &= (d-c)\mathbb{E}[Y_1] - B, \end{aligned} \tag{5.13}$$

$$\begin{aligned} S^2_{(c,d)} &= \mathbb{E}[R^2_{(c,d)}] - (\mathbb{E}[R_{(c,d)}])^2 \\ &= (-i)^2\phi''_{R_{(c,d)}}(0) - \left[(-i)\phi'_{R_{(c,d)}}(0)\right]^2 \\ &= (d-c)\mathbb{E}[Y_1^2]. \end{aligned} \tag{5.14}$$

Now, given that $U_1 \overset{\text{i.i.d.}}{\sim} \mathcal{U}(c,d)$, we have that, for $k \in \mathbb{N}$,

$$\begin{aligned} \mathbb{E}[U_1^{-k/\alpha}] &= \frac{1}{d-c}\int_c^d U^{-k/\alpha}\,\mathrm{d}U \\ &= \frac{1}{d-c}\left[\frac{\alpha}{\alpha-k}U^{-k/\alpha+1}\right]_c^d \\ &= \frac{1}{d-c}\frac{\alpha}{\alpha-k}\left(d^{\frac{\alpha-k}{\alpha}} - c^{\frac{\alpha-k}{\alpha}}\right). \end{aligned} \tag{5.15}$$

Finally, since W_1 is independent of U_1, then, for $k = 1, 2$,

$$\mathbb{E}[Y_1^k] = \mathbb{E}[W_1^k]\frac{1}{(d-c)}\frac{\alpha}{\alpha-k}\left(d^{\frac{\alpha-k}{\alpha}} - c^{\frac{\alpha-k}{\alpha}}\right). \tag{5.16}$$

Substituting (5.16) into the moment expressions (5.13) and (5.14) leads directly to the second part of the lemma. Observe that an alternative way of proving this second part is to use again the law of total expectation for $R_{(c,d)}$ directly, as was done in Riabiz et al. [2017b]. In any case, we need to assume $\mathbb{E}[W_1^2] < \infty$ for the second moment of Y_1 to be finite.

■

These expressions allow, in turn, to characterize the centring and scaling constants that are used in the CLT, respectively as

$$m_{(c,\infty)} := \lim_{d\to\infty} m_{(c,d)} = \mathbb{E}[W_1]\frac{\alpha}{1-\alpha}c^{\frac{\alpha-1}{\alpha}}, \tag{5.17}$$

$$S^2_{(c,\infty)} := \lim_{d\to\infty} S^2_{(c,d)} = \mathbb{E}[W_1^2]\frac{\alpha}{2-\alpha}c^{\frac{\alpha-2}{\alpha}}. \tag{5.18}$$

In principle, a similar limiting technique can be applied to the CF of $R_{(c,\infty)}$, where we can now state the identity

$$\phi_{R_{(c,\infty)}}(s) = \lim_{d\to\infty} \phi_{R_{(c,d)}}(s). \tag{5.19}$$

In fact, the CF of $\phi_{R_{(c,\infty)}}$, the expectation of a limit, is guaranteed to exist by the PSR; on the other the exchange of limit and expectation is now valid, thanks to the bounded convergence theorem, see [Rudin, 2006, Theorem 1.34]. This applies to CFs, given that their absolute value is uniformly bounded, see Lemma 9 in Appendix B.1.[4] We detail this procedure in the following section where an intermediate step (Taylor series expansion of $\phi_{Y_1}(s)$) is added to prove that the PSR residual is in the domain of attraction of the normal distribution (it satisfies the 'classical' CLT).

5.2 Asymptotic Normality of the PSR Residual

Even though it is composed as an infinite sum of random variables with finite moments, a careful study of its structure and CF shows that the residual $R_{(c,\infty)}$ is not Gaussian. However, the following CLT-like result states its asymptotic normality as $c \to \infty$. A first CLT for the special case $W_j = 1$ was presented in Lemke [2014], and the general case with random W_j was stated in Riabiz et al. [2017a]. Here we report the statement of the claim together with a complete proof, that was presented in Riabiz et al. [2018a] under milder moment conditions than those stated in Riabiz et al. [2017a].

Theorem 5.2.1. Let $R_{(c,\infty)}$, $m_{(c,\infty)}$ and $S^2_{(c,\infty)}$ be defined as in (5.3), (5.17) and (5.18), respectively. If $\mathbb{E}[|W_1|^3] < \infty$, then,

$$Z_{(c,\infty)} := \frac{R_{(c,\infty)} - m_{(c,\infty)}}{S_{(c,\infty)}} \xrightarrow[c\to\infty]{\mathcal{D}} Z, \tag{5.20}$$

where $Z \sim \mathcal{N}(0,1)$ and $\xrightarrow[c\to\infty]{\mathcal{D}}$ denotes convergence in distribution, as $c \to \infty$.

Proof: The proof is based on the Lévy's continuity theorem, see e.g. [Feller, 1966, p. 508]: we show that, for any fixed $s \in \mathbb{R}$, the CF of $Z_{(c,\infty)}$, $\phi_{Z_{(c,\infty)}}(s)$, converges to the CF of Z, as $c \to \infty$. We divide the proof in the following steps.
Step I. We express the CF $\phi_{Z_{(c,\infty)}}(s)$ in terms of the CF of the unnormalized residual

[4] An even simpler argument for the exchangeability of limit and expectation for the CF holds as follows. Given that the PSR (2.29) exists, the residual $R_{(c,\infty)}$ exists too, since it corresponds to the full series, except finitely many terms. This means that for fixed $c > 0$ the series in $R_{(c,d)}$ converges in some sense as $d \to \infty$. The weakest possible sense is convergence in distribution, which is equivalent to (5.19), via Lévy's continuity theorem, see e.g. [Feller, 1966, p. 508].

$\phi_{R_{(c,\infty)}}(s)$, and hence in terms of $\phi_{Y_1}(s)$, the CF of Y_1, defined in (5.6). Using the definition (5.20) of $Z_{(c,\infty)}$ and the rule for the CF of linear transformations of RVs, see Lemma 9 in Appendix B.1, we obtain

$$\phi_{Z_{(c,\infty)}}(s) = \exp\left(-i\frac{m_{(c,\infty)}}{S_{(c,\infty)}}s\right)\phi_{R_{(c,\infty)}}\left(\frac{s}{S_{(c,\infty)}}\right). \tag{5.21}$$

Now, taking the limit as in (5.19) and using expression (5.9), we have

$$\begin{aligned}\phi_{R_{(c,\infty)}}(s) &= \lim_{d\to\infty}\phi_{R_{(c,d)}}(s)\\ &= \lim_{d\to\infty}\exp\Big((d-c)\left(\phi_{Y_1}(s)-1\right)-iBs\Big),\end{aligned} \tag{5.22}$$

or, equivalently,

$$\log\phi_{R_{(c,\infty)}}(s) = \lim_{d\to\infty}\Big((d-c)\left(\phi_{Y_1}(s)-1\right)-iBs\Big). \tag{5.23}$$

Step II. Since we assumed that $\mathbb{E}[|W_1|^3] < \infty$, and $\mathbb{E}[|U_1|^{-3/\alpha}] = \mathbb{E}[U_1^{-3/\alpha}]$ is also finite and computed as in (5.15) we have that $\mathbb{E}[|Y_1|^3] = \mathbb{E}[|W_1|^3]\mathbb{E}[|U_1|^{-3/\alpha}] < \infty$. Thus we can write the following bound on the difference between $\phi_{Y_1}(s)$ and its Taylor expansion centred on 0 and truncated to the second order, see e.g. [Feller, 1966, p. 514],

$$\left|\phi_{Y_1}(s) - \sum_{k=0}^{2}\frac{i^k\mathbb{E}[Y_1^k]}{k!}s^k\right| \le \frac{\mathbb{E}[|Y_1|^3]}{3!}|s|^3.$$

Then, multiplying with $(d-c)$ the previous inequality, we obtain

$$\left|(d-c)\left(\phi_{Y_1}(s)-1\right) - (d-c)\sum_{k=1}^{2}\frac{i^k\mathbb{E}[Y_1^k]}{k!}s^k\right| \le (d-c)\frac{\mathbb{E}[|Y_1|^3]}{3!}|s|^3. \tag{5.24}$$

Step III. We recall that, from (5.13) and (5.14) in the proof of Lemma 2,

$$(d-c)\mathbb{E}[Y_1^k] = \begin{cases} m_{(c,d)} + B, & k=1,\\ S^2_{(c,d)}, & k=2,\end{cases} \tag{5.25}$$

while, from (5.16),

$$(d-c)\mathbb{E}[|Y_1|^3] = \mathbb{E}[|W_1|^3]\frac{\alpha}{\alpha-3}\left(d^{(\alpha-3)/\alpha} - c^{(\alpha-3)/\alpha}\right).$$

Hence (5.24) becomes

$$\left|(d-c)\left(\phi_{Y_1}(s)-1\right)-i(m_{(c,d)}+B)s+\frac{s^2S^2_{(c,d)}}{2}\right|\leq \mathbb{E}[|W_1|^3]\frac{\alpha}{\alpha-3}\left(d^{(\alpha-3)/\alpha}-c^{(\alpha-3)/\alpha}\right)\frac{|s|^3}{6}.$$

and taking the limit as $d\to\infty$ and using (5.23), we obtain

$$\left|\log\phi_{R_{(c,\infty)}}(s)-im_{(c,\infty)}s+\frac{s^2S^2_{(c,\infty)}}{2}\right|\leq \mathbb{E}[|W_1|^3]\frac{\alpha}{3-\alpha}c^{(\alpha-3)/\alpha}\frac{|s|^3}{6}. \tag{5.26}$$

Step IV. Finally, evaluating (5.26) at $s/S_{(c,\infty)}$ and using (5.21) and (5.18), we obtain

$$\left|\log\phi_{Z_{(c,\infty)}}(s)+\frac{s^2}{2}\right|\leq \frac{\mathbb{E}[|W_1|^3]}{\mathbb{E}[W_1^2]^{3/2}}\frac{\frac{\alpha}{3-\alpha}}{\left(\frac{\alpha}{2-\alpha}\right)^{3/2}}c^{-1/2}\frac{|s|^3}{6}.$$

Noting that the above right-hand side vanishes, as $c\to\infty$, we have

$$\log\phi_{Z_{(c,\infty)}}(s)\to-\frac{s^2}{2},$$

or, equivalently,

$$\lim_{c\to\infty}\phi_{Z_{(c,\infty)}}(s)=\exp\left(-\frac{s^2}{2}\right):=\phi_Z(s), \tag{5.27}$$

where $\phi_Z(s)$ is the CF of the standard Gaussian distribution, as required. ■

Remark 6. Even though it may seem intuitive that a classic central limit theorem (CLT) would hold for $R_{(c,\infty)}$, this cannot be applied directly. In fact, the summands involved in $R_{(c,\infty)}$ are non i.i.d., and their distribution depends on c and d, requiring care when Taylor-expanding and taking the limit of the CF $\phi_{Y_1}(s)$, as $c\to\infty$. Furthermore, an examination of the argument in the above proof indicates that the condition $\mathbb{E}[|W_1|^3]<\infty$ might be relaxed to $\mathbb{E}[W_1^2]<\infty$. But such an extension would require technical refinements that are beyond our present scope of using $W_j\sim\mathcal{N}(\mu_W,\sigma_W^2)$.

5.2.1 Gaussian Approximation of the Residual

Theorem 5.2.1 offers an asymptotic justification of the Gaussian approximation of the PSR residual $\hat{R}_{(c,\infty)}$

$$\hat{R}_{(c,\infty)}\sim\mathcal{N}\left(m_{(c,\infty)},S^2_{(c,\infty)}\right), \tag{5.28}$$

discussed in Sections 2.4.4 and 3.3.4 in the context of practical inference procedures. Given that $R_{(c,\infty)}$ converges in distribution to its Gaussian approximation as $c\to\infty$,

then by analogy with (5.1), and as shown at the bottom of Figure 5.1, we can introduce the following RV,

$$\hat{X} := X_{(0,c)} + \hat{R}_{(c,\infty)}, \tag{5.29}$$

that converges in distribution to $X \sim \mathcal{S}_\alpha(\sigma, \beta, \mu = 0)$, as $c \to \infty$.

Observe that Theorem 5.2.1 does not assume Gaussianity of the variables W_j. However, if this is the case, the following overall approximate conditionally Gaussian structure holds for the model, justifying our focus on $W_j \sim \mathcal{N}(\mu_W, \sigma_W^2)$ in the rest of this and the following Chapters.

5.2.2 Approximate Conditionally Gaussian Representation of the α-Stable Distribution

Assume $W_j \sim \mathcal{N}(\mu_W, \sigma_W^2)$, and assume that only the finite set $\{\Gamma_j \leq c\}$ is known, which is now a realistic assumption, as opposed to the infinite sequence $\{\Gamma_j\}_{j=1}^{\infty}$ required in the exact MaSMiN model (2.38a)- (2.38b). Then the α-stable distributed RV $X \sim \mathcal{S}_\alpha(\sigma, \beta, \mu = 0)$ has the approximate conditionally Gaussian representation

$$X | \{\Gamma_j \leq c\} \overset{\text{approx}}{\sim} \mathcal{N}\left(\mu_W M_{(0,c)} + m_{(c,\infty)},\ \sigma_W^2 S_{(0,c)}^2 + S_{(c,\infty)}^2\right), \tag{5.30}$$

where $M_{(0,c)}$ and $S_{(0,c)}^2$ are the RVs defined in (2.41) and (2.42), $m_{(c,\infty)}$ and $S_{(c,\infty)}^2$ have expressions (5.17) and (5.18).

Figure 5.2, inspired by Lemke [2014], compares kernel density estimations (KDEs) obtained with three different sampling methods: (i) $X_{(0,c)}$ is the approximation of X by the truncated PSR (5.2); (ii) $X_{(0,c)} + \hat{R}_{(c,\infty)}$ is the GAA for the PSR (5.29) (or the alternative generative scheme (5.30)); (iii) 'CMS' is the benchmark Chambers-Mallows-Stuck method for generating stable RVs as in Theorem 2.1.1 from Chapter 2. The figure indicates that adding the Gaussian approximation of the residual to $X_{(0,c)}$ results in an approximate distribution that is closer to the true stable law than that obtained by simple truncation of the PSR. Furthermore, the convergence of $X_{(0,c)}$ to X might be very slow, especially for certain configurations of the distribution parameters.

This serves to illustrate the power of Theorem 5.2.1: the inference methods valid for the exact PSR can be used for its approximation (5.29), whose quality is controlled directly by the truncation parameter c. In fact, the inference schemes by Lemke and Godsill [2012, 2014, 2015]; Lemke et al. [2015] referenced in Section 3.3.4 were based on (5.30). However, in these references, there is no quantitative measure of *how good*

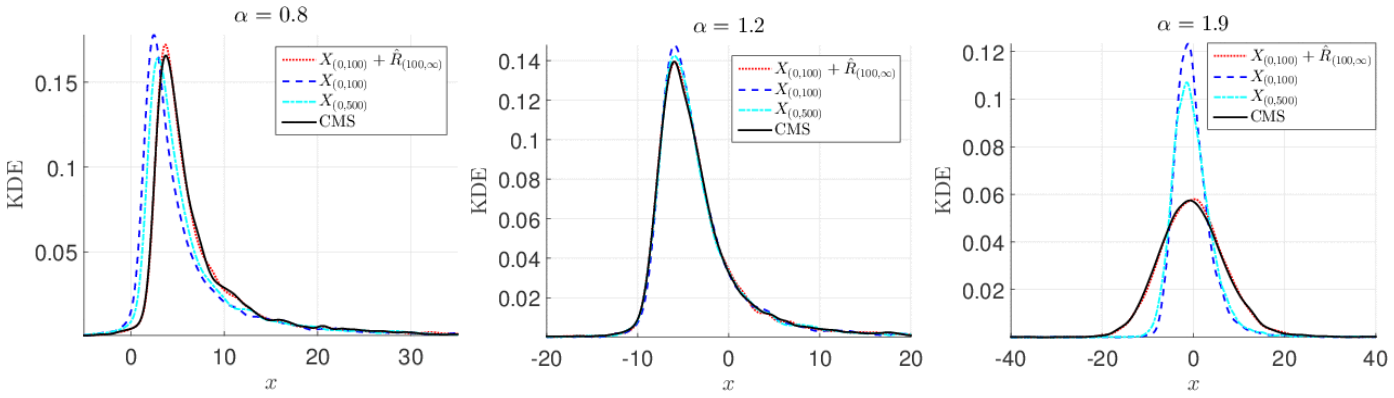


Figure 5.2 Kernel density estimates over 10^4 RVs generated with different schemes: $X_{(0,c)}$ is the truncated PSR (5.2), $X_{(0,c)} + \hat{R}_{(c,\infty)}$ is our proposed approximation for the PSR (5.29), CMS is the benchmark Chambers-Mallows-Stuck method for generating stable RVs. We compare results for three values of tail parameter $\alpha = 0.8$ (left), $\alpha = 1.2$ (centre), $\alpha = 1.9$ (right), and for two values of truncation parameter $c = 100$ and $c = 500$.

the approximation is and how it might vary with the truncation parameter c and the distribution parameters $\{\alpha, \beta, \sigma\}$, or, alternatively, $\{\alpha, \mu_W, \sigma_W\}$ in our Gaussian scenario for W_j. These issues are addresses in the rest of this and the following two Chapters.

5.2.3 Selection of the Truncation Parameter c

In order to quantify the approximation error in the representation $X_{(0,c)} + \hat{R}_{(c,\infty)}$, and also in order to be able to choose appropriate values for the truncation parameter c, some considerations should be kept in mind:

- the distribution of $R_{(c,\infty)}$ matches better that of $\hat{R}_{(c,\infty)}$ when c is increased, according to Theorem 5.2.1. This is a justification for choosing larger values of c;
- the computational complexity of the approximate conditionally Gaussian model (5.30) increases when c is increased. In fact, the expected cardinality of the set of latent RVs $\{\Gamma_j < c\}$ that we need to generate to compute $M_{(0,c)}$ and $S^2_{(0,c)}$ is c. This by contrast would make us wish to choose small values for c;
- finally, even if $R_{(c,\infty)}$ is far from being Gaussian because c is not large enough, its contribution to the full PSR X might be relatively 'small', when compared to $X_{(0,c)}$.

We will consider a 'good' choice of c as one that makes the distribution of $\hat{X}$ close to that of X, hence giving more weight to the last of the above points. Quantifying the

distance between $X$and $\hat{X}$ involves computing the distance between $R_{(c,\infty)}$ and $\hat{R}_{(c,\infty)}$ so we proceed by first estimating how far the PSR residual is from the corresponding Gaussian, for finite c.

5.3 Divergence Measures

In view of the above discussion, our aim is to provide accurate bounds that can guide the practical choice of the truncation parameter c, given the distribution parameters. The main tools that we employ to derive such bounds are based on classical Fourier-analytic techniques, that convert the notion of convergence between CFs, used in the proof of Theorem 5.2.1, to convergence[5] between CDFs and PDFs. We recur to these tools because the expressions of the desired CDFs and PDFs are not available in closed form, due to the intractability of the stable law.[6]

We present our findings, following their chronological development. We first performed Monte Carlo and numerical simulations, in order to establish the intuition for the closed form bounds. These were then found by a study of the quantities involved in the Fourier-analytic tools mentioned above. We underscore that, while the considerations stemming from the Monte Carlo simulations are valid in any parameter regime, the sharp closed form bounds (and their numerical simulations) that we provide in this thesis are related to distances between CDFs, and for the special case of symmetric stable laws ($\beta = 0$). We leave to future developments the extension of our studies to stable distributions with an arbitrary skewness parameters, and the formulation of analytic bounds for distances between PDFs, for which we only provide numerical results here.

5.3.1 Monte Carlo Simulations of the CDF Divergence

Suppose S and T are RVs with CDFs $F_S(x)$ and $F_T(x)$, $s \in \mathbb{R}$, respectively. The divergence measure between CDFs that we consider is

$$\Delta(S,T) := \sup_{x\in\mathbb{R}} |F_S(x) - F_T(x)|\,,$$

also known as the Kolmogorov distance between the distributions of S and T, see e.g. Rachev et al. [2013].

[5]We interchange the terms 'divergence', 'convergence', and 'distance'. The first term is the more generic, while the second one is justified by the fact that we look at measures of divergence that

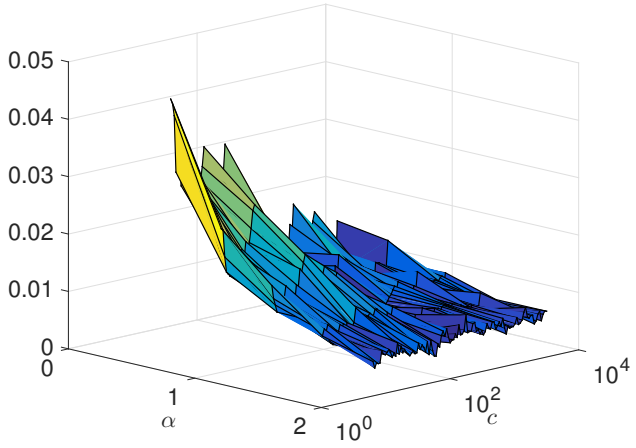


Figure 5.3 Kolmogorov-Smirnov test statistic for $Z_{(c,d)}$, with $N = 10^4$, $d = 10^5$, $\mu_W = 1, \sigma_W = 1$, on a grid of c and α values.

Here we present the experimental study[7] of the Kolmogorov distance between the CDF of the standard Gaussian RV Z and the the empirical CDF of the standardized version of the residual $R_{(c,d)}$, namely[8]

$$Z_{(c,d)} := \frac{R_{(c,d)} - m_{(c,d)}}{S_{(c,d)}}, \tag{5.31}$$

where all the quantities in the definition of $Z_{(c,d)}$ have been defined in the previous two sections. Given a sample of realizations of the RV $Z_{(c,d)}$, denoted $\boldsymbol{Z}_{(c,d)} := [Z^{(1)}_{(c,d)}, \dots, Z^{(N)}_{(c,d)}]$ in the following, the empirical CDF of $Z_{(c,d)}$ is

$$\hat{F}_{Z_{(c,d)}}(x) := \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}(Z^{(n)}_{(c,d)} \in (-\infty, x)), \tag{5.32}$$

where $\mathbb{1}(\cdot)$ is the indicator function. Then, we look at the Kolmogorov-Smirnov statistic, defined as

$$\hat{\Delta}(Z_{(c,d)}, Z) := \sup_{x \in \mathbb{R}} \left| \hat{F}_{Z_{(c,d)}}(x) - F_Z(x) \right|, \tag{5.33}$$

vanish asymptotically with c; finally, the third term is appropriate because the divergences that we consider are symmetric and do satisfy the triangle inequality.

[6]For this reason it is also not possible to make direct use of common divergence measures, such as Kullback-Leibler divergence (KLD) and α-divergence, see Minka [2005].

[7]These results were first given in Riabiz et al. [2017b].

[8]Observe that, from a simulation perspective, if we want to have control on the LVs $\{\Gamma_j\}$, this is the only quantity that can be generated. In fact, even if we can formally express $Z_{(c,\infty)} = (X - X_{(0,c)} - m_{(c,\infty)})/S_{(c,\infty)}$, where all the quantities involved can be simulated/computed, we cannot guarantee that the LVs in X (drawn using the CMS method) are the same as those in $X_{(0,c)}$ (drawn summing a few terms of the PSR). To make $Z_{(c,d)}$ be as close as possible to $Z_{(c,\infty)}$, we set $d \gg 1$, and as large as computationally affordable.

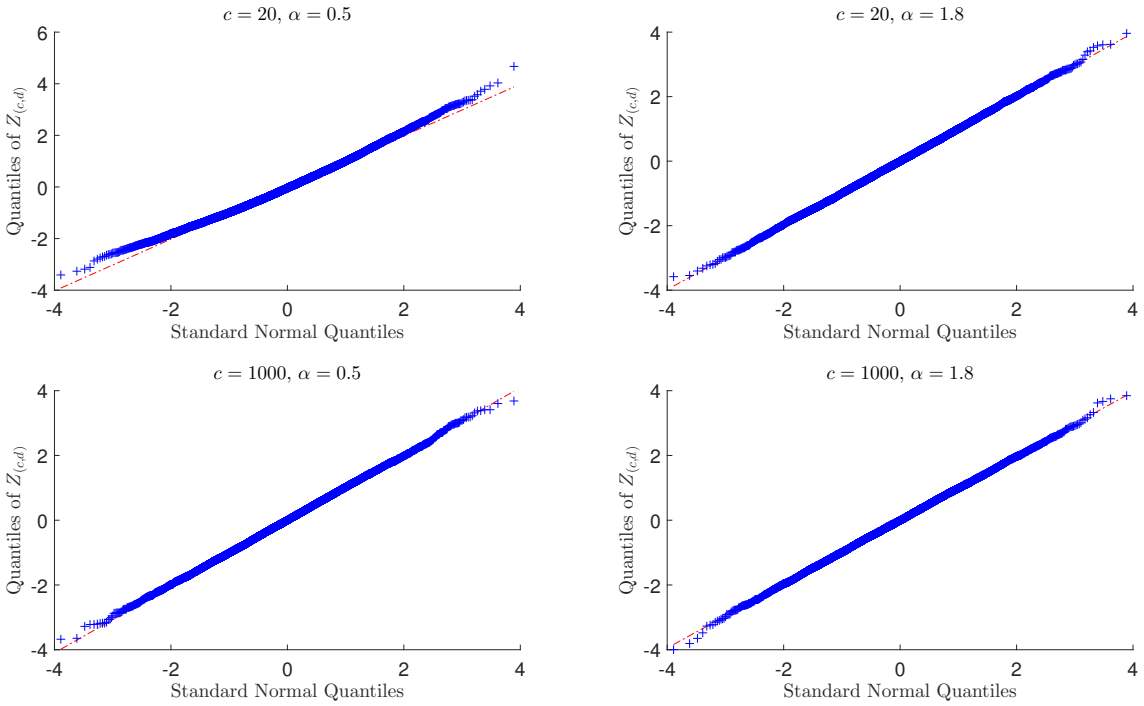


Figure 5.4 Comparison of the quantiles of $Z_{(c,d)}$, with those of a standard Gaussian $d = 10^5, N = 10^4, \mu_W = 1, \sigma_W = 1$ for $c \in \{20, 1000\}$, $\alpha \in \{0.5, 1.8\}$.

and used in the homonym statistical test for normality, see Massey [1951]: large values put evidence in favour of large distance between the distribution of $Z_{(c,d)}$ and a standard Gaussian, and vice-versa.

In our experiment, we work with a sample $\boldsymbol{Z}_{(c,d)}$ of size $N = 10^4$; we set $d = 10^5$ and we increase c on a logarithmic scale of values between 100 and 5×10^3. Furthermore we use $\mu_W = 1$, $\sigma_W = 1$, and we repeat the computation of $\hat{\Delta}(Z_{(c,d)}, Z)$ for a grid of values of $\alpha \in \{0.3, 0.7, 1.1, 1.5, 1.9\}$, corresponding, respectively, to $\beta = \{0.75, 0.82, 0.86, 0.9, 0.92\}$, and $\sigma = \{1.52, 1.66, 1.90, 2.4, 4.8\}$, through the transformations (2.34a)-(2.34b). The result is reported in Fig. 5.3. We remark that the surface plot shown is non-smooth because the variables $Z^{(n)}_{(c,d)}$ in the empirical CDF (5.32) are random, so that the supremum in (5.33) is not deterministically decaying as c or α increase. Moreover, the coarse grid that was used for the variable α introduce some further apparent non-smoothness. However, it is possible to do the following two observations: (i) as we increase c for a fixed value of α, the test statistic decreases, supporting the hypothesis that the sample comes from a Gaussian distribution, as expected from Theorem 5.2.1; (ii) for a fixed value of c, the test indicates that Gaussianity increases as α tends to 2; this is not surprising given that the α-stable distribution approaches the Gaussian, α tends to 2.

To support the observations obtained from the Kolmogorov-Smirnov test statistics, in Fig.5.4 we show Q-Q plots of the quantiles of the sample distribution of $Z_{(c,d)}$, compared to the standard Gaussian quantiles. We consider a 'very' heavy-tailed stable distribution ($\alpha = 0.5$) and a 'less' heavy tailed one ($\alpha = 1.8$), as well as two values for c, a lower one ($c = 20$) and a higher one ($c = 10^3$). As expected, the normalized residuals $Z_{(c,d)}$ are still heavy tailed when $\alpha = 0.5, c = 20$ and the Gaussianity is better met when c is increased. On the other hand, when $\alpha = 1.8$, the quantiles of the empirical distribution resemble the Gaussian already when $c = 20$.

When d is large, $\hat{\Delta}(Z_{(c,d)}, Z)$ gives interesting insight on the departure from Gaussianity of $Z_{(c,\infty)}$ for finite values of c, and as a function of the distribution parameters. However, the test statistic is a RV and, as such, its realizations do not give repeatable bounds for $\Delta(Z_{(c,\infty)}, Z)$; moreover this procedure is not practical when one wants to obtain bounds for an arbitrary/large number of parameter values. This justified our subsequent work Riabiz et al. [2017c], Riabiz et al. [2018b], Riabiz et al. [2018a], presented in the following.

5.3.2 CDF Divergence and Smoothing Lemma

Let S and T, $F_S(x)$ and $F_T(x)$, $x \in \mathbb{R}$, be defined as in the previous section. Suppose also that $\phi_S(s)$ and $\phi_T(s)$, $s \in \mathbb{R}$, are the respective CFs, and assume that $\mathbb{E}[S] = \mathbb{E}[T] = 0$. Furthermore, assume that $F_T(x)$ has derivative $p_T(x)$ such that $|p_T(x)| \leq m < \infty$, $\forall x \in \mathbb{R}$. Then, Esséen's smoothing lemma [Feller, 1966, Lemma 2, p. 538] states that, for any $\Theta > 0$,

$$\Delta(S,T) \leq \frac{1}{\pi}\int_{-\Theta}^{\Theta} \frac{|\phi_S(s) - \phi_T(s)|}{|s|}\,\mathrm{d}s + \frac{24m}{\pi\Theta} := I(S,T) \tag{5.34}$$

$$\overset{\Theta\to\infty}{\longrightarrow} \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{|\phi_S(s) - \phi_T(s)|}{|s|}\,\mathrm{d}s := \bar{I}(S,T), \tag{5.35}$$

where (5.35) is meaningful only if the improper integral converges, otherwise the smoothed version (5.34) should be considered. Notice that the integrands in the above expressions have a removable singularity at $s = 0$, thanks to the assumption of vanishing expectations of S and T. However, such an assumption can be relaxed, if either (5.34) or (5.35) are finite.

As before, let $c \geq 0$ be the value of the truncation parameter of the PSR. In Chapter 6 we use the smoothing lemma to first investigate the convergence of the PSR standardized residual to the Gaussian distribution, by deriving an upper bound

for the distance,

$$\Delta(Z_{(c,\infty)}, Z) := \sup_{x\in\mathbb{R}} \left| F_{Z_{(c,\infty)}}(x) - F_Z(x) \right|, \tag{5.36}$$

where $F_{Z_{(c,\infty)}}(x)$ and $F_Z(x)$ denote the CDF of $Z_{(c,\infty)}$ and the standard normal CDF, respectively.

We then use this result in Chapter 7 to further bound the Kolmogorov distance between the approximated stable law with Gaussian approximation of the PSR residual, $\hat{X}$, and the 'exact' stable law, X,

$$\Delta(X, \hat{X}) := \sup_{x\in\mathbb{R}} \left| F_X(x) - F_{\hat{X}}(x) \right|, \tag{5.37}$$

where $F_X(x)$ and $F_{\hat{X}}(x)$ are the CDFs of X and $\hat{X}$, respectively. As mentioned in Section 2.4.4, an alternative approximation of the PSR consists in just truncating the series, which has the following Kolmogorov distance from the full PSR X,

$$\Delta(X, X_{(0,c)}) := \sup_{x\in\mathbb{R}} \left| F_X(x) - F_{X_{(0,c)}}(x) \right|, \tag{5.38}$$

where $F_{X_{(0,c)}}(x)$ is the CDF of $X_{(0,c)}$ (5.2). In detail, we will analyse the effect of our Gaussian approximation of the residual (5.28), by comparing the two bounds on $\Delta(X, \hat{X})$ and $\Delta(X, X_{(0,c)})$.

5.3.3 PDF Divergence and Continuity Theorem for Densities

Suppose that S and T are RVs as above, with PDFs $f_S(x)$ ad $f_T(x)$. If $\phi_S(s)$ and $\phi_T(s)$ are integrable functions, then we can use the continuity theorem for densities, see [Feller, 1966, p.510], to uniformly bound the distance between their PDFs

$$\delta(S, T) := |f_S(x) - f_T(x)| \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} |\phi_S(s) - \phi_T(s)| \, \mathrm{d}s := i(S, T). \tag{5.39}$$

In Chapter 6 we report numerical simulations of $i(Z_{(c,\infty)}, Z)$, the bound between the PDF of $Z_{(c,\infty)}$ and Z.

In order to bound (5.36), (5.37) and (5.38), we need explicit expressions for the various CFs of interest. These are derived in the following, and are the second main contribution of this Chapter.

Table 5.1 Summary of the logarithms of the CF expressions derived and used. Recall that $\alpha \in (0,2)$, $\alpha \neq 1$, and $c > 0$.

Distribution W_j	Skewness	RV	$\log(\phi(s))$ or $\log(\psi(w))$ or $\log(\omega(u))$, with w and u as in (5.57)	Equations
S.t. (2.31) and (5.46)	$\beta \in [-1,1]$	X	$-\sigma^\alpha \lvert s\rvert^\alpha \left\{1 - i\beta\,\mathrm{sgn}(s)\tan\frac{\pi\alpha}{2}\right\}$	(2.3), (2.34a)-(2.34b)
		$R_{(c,d)}$	$ism_{(c,d)} - s^2 S^2_{(c,d)}/2 + \sum_{k=3}^{\infty} r_k s^k$	(5.11), (5.48), (5.49)
		$R_{(c,\infty)}$	$ism_{(c,\infty)} - s^2 S^2_{(c,\infty)}/2 + \sum_{k=3}^{\infty} \bar{r}_k s^k$	(5.50), (5.51), (5.18)
		$Z_{(c,\infty)}$	$-s^2/2 + \sum_{k=3}^{\infty} \bar{z}_k s^k$	(5.44) (5.45)
		Z	$-s^2/2$ or $-cw/\eta$	(5.27), (5.63)
		$\hat{R}_{(c,\infty)}$	$-s^2 S^2_{(c,\infty)}/2 + ism_{(c,\infty)}$	(5.17), (5.18), (5.43)
$\mathcal{N}(\mu_W, \sigma_W^2)$	$\beta \in [-1,1]$	$R_{(c,d)}$	$c - d + \alpha \int_{d^{-1/\alpha}}^{c^{-1/\alpha}} e^{(ist\mu_W - \sigma_W^2 s^2 t^2/2)} t^{-\alpha-1}\,\mathrm{d}t - isB$	(5.53), (5.8)
			$\alpha \int_{d^{-1/\alpha}}^{c^{-1/\alpha}} \left(e^{(ist\mu_W - \sigma_W^2 s^2 t^2/2)} - 1 - is\mu_W t\right) t^{-\alpha-1}\,\mathrm{d}t - is\mu_W \frac{\alpha}{\alpha-1} c^{\frac{\alpha-1}{\alpha}}$	(5.54)
		$R_{(c,\infty)}$	$\alpha \int_{0}^{c^{-1/\alpha}} \left(e^{(ist\mu_W - \sigma_W^2 s^2 t^2/2)} - 1 - is\mu_W t\right) t^{-\alpha-1}\,\mathrm{d}t - is\mu_W \frac{\alpha}{\alpha-1} c^{\frac{\alpha-1}{\alpha}}$	(5.52)
		$X_{(0,c)}$	$-c + \int_{c^{-1/\alpha}}^{\infty} e^{(ist\mu_W - \sigma_W^2 s^2 t^2/2)} t^{-\alpha-1}\,\mathrm{d}t$	(5.55)
		X	$\alpha \int_0^\infty \left(e^{(ist\mu_W - \sigma_W^2 s^2 t^2/2)} - 1\right) t^{-\alpha-1}\,\mathrm{d}t$	(5.56)
$\mathcal{N}(0, \sigma_W^2)$	$\beta = 0$	$R_{(c,\infty)}$	$c\Big(1 - \exp(-u) - u^a \gamma\,(1-a,u)\Big)$	(5.67)
		$Z_{(c,\infty)}$	$c\Big(1 - \exp(-w) - w^a \gamma\,(1-a,w)\Big)$	(5.58)
		$X_{(0,c)}$	$-c(1 - \exp(-u) + u^a \Gamma(1-a,u))$	(5.68)
		$\hat{R}_{(c,\infty)}$	$-cu/\eta$	(5.70)
		$\hat{X}$	$-c(1 - \exp(-u) + u^a \Gamma(1-a,u) + u/\eta)$	(5.71)
		X	$\exp(-\sigma_W^\alpha 2^{-a}\Gamma(1-a)\lvert s\rvert^\alpha)$	(5.69)

5.4 Characteristic Function Expressions

Recall that $X \sim \mathcal{S}_\alpha(\sigma, \beta, \mu = 0)$ has CF given in (2.3), and that we approximate it with $\hat{X}$ (5.29). Given that $R_{(c,\infty)}$ and $\hat{R}_{(c,\infty)}$ are independent, of $X_{(0,c)}$, using Lemma 10 in Appendix B.1 we immediately have the following relations between CFs

$$\phi_X(s) = \phi_{X_{(0,c)}}(s)\phi_{R_{(c,\infty)}}(s), \tag{5.40}$$

$$\phi_{\hat{X}}(s) = \phi_{X_{(0,c)}}(s)\phi_{\hat{R}_{(c,\infty)}}(s). \tag{5.41}$$

Furthermore, the CFs for the true and approximated residuals are obtained as affine transformations of their normalised counterparts by a simple change of variable (see Lemma 9 in Appendix B.1)

$$\phi_{R_{(c,\infty)}}(s) \;\; = \phi_{Z_{(c,\infty)}}(S_{(c,\infty)}s)\exp\left(ism_{(c,\infty)}\right), \tag{5.42}$$

$$\phi_{\hat{R}_{(c,\infty)}}(s) = \phi_Z(S_{(c,\infty)}s)\exp\left(ism_{(c,\infty)}\right), \tag{5.43}$$

with $m_{(c,\infty)}$ and $S_{(c,\infty)}$ their mean and standard deviation, as in (5.17) and (5.18), and $\phi_Z(s)$ the CF of the standard normal distribution, as in (5.27).

From these relations between CFs, it is clear that, in order to use the smoothing lemmas for (5.36) and (5.37) we need to obtain either analytic expressions or bounds for $\phi_{Z_{(c,\infty)}}(s)$, $\phi_{R_{(c,\infty)}}(s)$ and $\phi_{X_{(0,c)}}(s)$. These are derived in the following two subsections, in the case $W_1 \sim \mathcal{N}(\mu_W, \sigma_W^2)$, with special focus on the instances $\mu_W = 0$, corresponding to the symmetric stable distribution $\beta = 0$. The results were first presented in Riabiz et al. [2017a] and Riabiz et al. [2018a]. For easy reference, our findings are summarized in Table 5.1.

5.4.1 CF Expressions when $W_1 \sim \mathcal{N}(\mu_W, \sigma_W^2)$

Series Expansion of $\phi_{Z_{(c,\infty)}}(s)$

If W_1 is Gaussian, then it is possible to write the CF of the PSR residual in terms of an infinite series.

Lemma 3. Let $Z_{(c,\infty)}$ be defined as in (5.20), and let $W_1 \sim \mathcal{N}(\mu_W, \sigma_W^2)$. Then, for $s \in \mathbb{R}$ and $c > 0$,

$$\phi_{Z_{(c,\infty)}}(s) = \exp\left(-\frac{s^2}{2} + \sum_{k=3}^{\infty} \bar{z}_k s^k\right), \tag{5.44}$$

where

$$\bar{z}_k := \frac{i^k}{k!} \frac{\mathbb{E}[W_1^k]^{\frac{\alpha}{k-\alpha}}}{\left(\mathbb{E}[W_1^2]^{\frac{\alpha}{2-\alpha}}\right)^{k/2}} c^{1-k/2}, \quad k \geq 3. \tag{5.45}$$

Proof: As in the proof of Theorem 5.2.1, we note that that the CF of $R_{(c,\infty)}$ (5.22) can be expressed in terms of the CF of $Y_1 = W_1 U_1^{-1/\alpha}$ defined in (5.6). Thus, we begin by expanding $\phi_{Y_1}(s)$ as a Taylor series.

Suppose $W_1 \sim \mathcal{N}(\mu_W, \sigma_W^2)$. Using (5.16) and the well-known formula for the moments of the normal distribution, it is possible to check that

$$\limsup_{k\to\infty} \frac{1}{k} |\mathbb{E}[Y_1^k]|^{1/k} < \infty, \tag{5.46}$$

as proved in Appendix B.2. Therefore, $\phi_{Y_1}(s)$ is analytic[9] around $s = 0$ and admits the Taylor expansion [Feller, 1966, p. 514]

$$\phi_{Y_1}(s) = 1 + i\mathbb{E}[Y_1]s - \frac{1}{2}\mathbb{E}[Y_1^2]s^2 + \sum_{k=3}^{\infty} \frac{i^k}{k!}\mathbb{E}[Y_1^k]s^k. \tag{5.47}$$

Then, from (5.9), (5.16) and (5.25), for $c > 0$, we have

$$\phi_{R_{(c,d)}}(s) = \exp\left(ism_{(c,d)} - \frac{s^2 S_{(c,d)}^2}{2} + \sum_{k=3}^{\infty} r_k s^k \right), \tag{5.48}$$

where

$$r_k := \frac{i^k}{k!}\mathbb{E}[W_1^k] \frac{\alpha}{\alpha - k} \left(d^{(\alpha-k)/\alpha} - c^{(\alpha-k)/\alpha} \right). \tag{5.49}$$

Recalling (5.19), we now want to take the limit as $d \to \infty$ of (5.48). Given that the series in this expression converges uniformly, it is possible to exchange the limit and series operators (see Appendix B.3 for the details), obtaining

$$\phi_{R_{(c,\infty)}}(s) = \lim_{d\to\infty} \phi_{R_{(c,d)}}(s) = \exp\left(ism_{(c,\infty)} - \frac{s^2 S_{(c,\infty)}^2}{2} + \sum_{k=3}^{\infty} \bar{r}_k s^k \right), \tag{5.50}$$

[9] The radius of convergence R of the power series (5.47) around $s = 0$ is such that

$$R^{-1} = \limsup_{k\to\infty} \left| \frac{i^k}{k!}\mathbb{E}[Y_1^k] \right|^{1/k} = \limsup_{k\to\infty} \frac{e}{k} |\mathbb{E}[Y_1^k]|^{1/k} < \infty,$$

where the second equality holds because Stirling's formula $n! \sim \sqrt{2\pi n}(n/e)^n$ implies that $(n!)^{1/n} \sim (n/e)$, see Dutka [1991], while the finiteness of the limit holds thanks to (5.46).

where

$$\bar{r}_k = \frac{i^k}{k!}\mathbb{E}[W_1^k]\frac{\alpha}{k-\alpha}c^{(\alpha-k)/\alpha}, \quad k \geq 3. \tag{5.51}$$

Hence, from (5.21) and (5.18),

$$\phi_{Z_{(c,\infty)}}(s) = \exp\left(-\frac{s^2}{2} + \lim_{d\to\infty}\sum_{k=3}^{\infty} z_k s^k\right) = \exp\left(-\frac{s^2}{2} + \sum_{k=3}^{\infty}\bar{z}_k s^k\right),$$

where the term-by-term limit can be justified as before,

$$\begin{aligned} z_k &:= \frac{i^k}{k!}\frac{(d-c)\mathbb{E}[Y_1^k]}{\left((d-c)\mathbb{E}[Y_1^2]\right)^{k/2}} \\ &= \frac{i^k}{k!}\frac{\mathbb{E}[W_1^k]\frac{\alpha}{\alpha-k}\left(d^{(\alpha-k)/\alpha} - c^{(\alpha-k)/\alpha}\right)}{\left(\mathbb{E}[W_1^2]\frac{\alpha}{\alpha-2}(d^{(\alpha-2)/\alpha} - c^{(\alpha-2)/\alpha})\right)^{k/2}}, \end{aligned}$$

and

$$\bar{z}_k := \frac{i^k}{k!}\frac{\mathbb{E}[W_1^k]\frac{\alpha}{k-\alpha}}{\left(\mathbb{E}[W_1^2]\frac{\alpha}{2-\alpha}\right)^{k/2}}c^{1-k/2}.$$

■

An examination of the proof shows that, while establishing the lemma, we also obtain the CF of the unnormalized residuals $R_{(c,d)}$ and $R_{(c,\infty)}$, as reported in Table 5.1. Moreover, it is also clear from the proof that the results hold not only in the case $W_1 \sim \mathcal{N}(\mu_W, \sigma_W^2)$, but more generally for any distribution on W_1 that satisfies the condition (5.46). Finally, we remark that the condition $c > 0$ in the statement of the Lemma is necessary, because $\mathbb{E}[Y_1^k]$ diverges for $c \downarrow 0$, and the condition (5.46) is then not satisfied.

Integral Expression for $\phi_{R_{(c,\infty)}}(s)$

Alternatively to the series expansion above, we can perform direct computations with W_1 normally distributed, obtaining the following integral expressions for the CF of the residual.

Lemma 4. Let $R_{(c,\infty)}$ be defined as in (5.3), and let $W_1 \sim \mathcal{N}(\mu_W, \sigma_W^2)$. Then, for $s \in \mathbb{R}$,

$$\log\phi_{R_{(c,\infty)}}(s) = \alpha\int_0^{c^{-1/\alpha}}\left(e^{(ist\mu_W - \sigma_W^2 s^2 t^2/2)} - 1 - is\mu_W t\right)t^{-\alpha-1}\,\mathrm{d}t - is\mu_W\frac{\alpha}{\alpha-1}c^{\frac{\alpha-1}{\alpha}}. \tag{5.52}$$

Proof: For convenience, we express the variable Y_1 in (5.6) as $Y_1 = W_1 t(U_1)$, where $t(x) := x^{-1/\alpha}$. Then we can write the CF of Y_1 directly as

$$\begin{aligned}\phi_{Y_1}(s) = \mathbb{E}[e^{isY_1}] &= \mathbb{E}[e^{isW_1 t(U_1)}] \\ &= \int_c^d \int_{\mathbb{R}} e^{iswt(u)} p_{(W_1,U_1)}(w,u)\,\mathrm{d}w\,\mathrm{d}u \\ &= \int_c^d \int_{\mathbb{R}} e^{iswt(u)} p_{W_1}(w) p_{U_1}(u)\,\mathrm{d}u\,\mathrm{d}u \\ &= \int_c^d \frac{\phi_{W_1}(st(u))}{d-c}\,\mathrm{d}u,\end{aligned}$$

where $p_{(W_1,U_1)}(w,u)$ is the joint density of the random vector (W_1, U_1) and $p_{W_1}(w)$ and $p_{U_1}(u)$ are the respective marginals, while $\phi_{W_1}(s)$ is the CF of a Gaussian $W_1 \sim \mathcal{N}(\mu_W, \sigma_W^2)$. Now, substituting the expression of $\phi_{W_1}(s)$,

$$\begin{aligned}\phi_{Y_1}(s) &= \int_c^d \frac{\exp(ist(u)\mu_W - \sigma_W^2 s^2 t(u)^2/2)}{d-c}\,\mathrm{d}u \\ &= \frac{1}{d-c}\int_{t(c)}^{t(d)} \frac{\exp(ist\mu_W - \sigma_W^2 s^2 t^2/2)}{\mathrm{d}t/\,\mathrm{d}u}\,\mathrm{d}t \\ &= \frac{\alpha}{d-c}\int_{d^{-1/\alpha}}^{c^{-1/\alpha}} \exp(ist\mu_W - \sigma_W^2 s^2 t^2/2) t^{-\alpha-1}\,\mathrm{d}t,\end{aligned}$$

and recalling (5.9), we have

$$\log(\phi_{R_{(c,d)}}(s)) = c - d + \alpha \int_{d^{-1/\alpha}}^{c^{-1/\alpha}} e^{(ist\mu_W - \sigma_W^2 s^2 t^2/2)} t^{-\alpha-1}\,\mathrm{d}t - isB. \tag{5.53}$$

By simple algebra, it is possible to re-write

$$\log(\phi_{R_{(c,d)}}(s)) = \alpha \int_{d^{-1/\alpha}}^{c^{-1/\alpha}} \left(e^{(ist\mu_W - \sigma_W^2 s^2 t^2/2)} - 1 - is\mu_W t\right) t^{-\alpha-1}\,\mathrm{d}t - is\mu_W \frac{\alpha}{\alpha-1} c^{\frac{\alpha-1}{\alpha}}, \tag{5.54}$$

and taking the limit $d \to \infty$,

$$\log(\phi_{R_{(c,\infty)}}(s)) = \alpha \int_0^{c^{-1/\alpha}} \left(e^{(ist\mu_W - \sigma_W^2 s^2 t^2/2)} - 1 - is\mu_W t\right) t^{-\alpha-1}\,\mathrm{d}t - is\mu_W \frac{\alpha}{\alpha-1} c^{\frac{\alpha-1}{\alpha}}.$$

■

Integral Expressions for $\phi_{X_{(0,c)}}(s)$ and $\phi_X(s)$

Notice that we have not made any assumptions about on the finiteness of the moments of $R_{(c,d)}$ or Y_1 in the proof of Lemma 4. We may then use the expression of the CF of $R_{(c,d)}$ with $c \downarrow 0$ to obtain an analogous expression for the CF of the truncated

PSR $X_{(0,c)}$. Observe that, in this case we do not have the term iBs appearing in (5.53), because all the centering constants have been included in the PSR residual $R_{(c,d)}$, see (5.4). Hence we have

$$\begin{aligned}\log(\phi_{X_{(0,c)}}(s)) &= -c + \int_{c^{-1/\alpha}}^{\infty} e^{(ist\mu_W - \sigma_W^2 s^2 t^2/2)} t^{-\alpha-1}\,\mathrm{d}t \\ &= \alpha \int_{c^{-1/\alpha}}^{\infty} \left(e^{(ist\mu_W - \sigma_W^2 s^2 t^2/2)} - 1\right) t^{-\alpha-1}\,\mathrm{d}t.\end{aligned} \tag{5.55}$$

Finally we obtain an integral expression for the CF of the full PSR X

$$\begin{aligned}\log(\phi_X(s)) &= \log(\phi_{X_{(0,c)}}(s)) + \log(\phi_{R_{(c,\infty)}}(s)) \\ &= \alpha \int_0^{\infty} \left(e^{(ist\mu_W - \sigma_W^2 s^2 t^2/2)} - 1\right) t^{-\alpha-1}\,\mathrm{d}t,\end{aligned} \tag{5.56}$$

as reported in Table 5.1.

5.4.2 CF Expressions when $W_1 \sim \mathcal{N}(0, \sigma_W^2)$

Analytic Expression for $\phi_{Z_{(c,\infty)}}(s)$

In the specific case of central W_1, we are able to provide more explicit expressions for the CF of the normalized residual $Z_{(c,\infty)}$, as reported in Lemma 5. This expression was first derived in Riabiz et al. [2017a], by summing the series (5.44), but the same result can be obtained based on the integral representation (5.52) followed by standardization, as presented in Riabiz et al. [2018a]. We report both proofs for completeness.

Lemma 5. Suppose $W_1 \sim \mathcal{N}(0, \sigma_W^2)$, and denote,

$$a := \frac{\alpha}{2}, \quad \eta := \frac{1-a}{a}, \quad w := \frac{\eta s^2}{2c}, \quad u := wS^2_{(c,\infty)}, \tag{5.57}$$

for $\alpha \in (0,2), \alpha \neq 1$, where $S^2_{(c,\infty)}$ is given as in (5.18), with $\mathbb{E}[W_1^2] = \sigma_W^2$. Then,

$$\begin{aligned}\phi_{Z_{(c,\infty)}}(s) &= \psi_{Z_{(c,\infty)}}(w) \\ &= \exp\left(c(1 - e^{-w} - w^a \gamma\,(1-a, w))\right),\end{aligned} \tag{5.58}$$

where $\gamma(s,x)$ is the lower incomplete gamma function (B.10).

Proof: Summing the series. Let us define

$$\xi_{c,\alpha}(s) := \sum_{k=3}^{\infty} \bar{z}_k s^k \tag{5.59}$$

to be the series in (5.44), where the coefficients $\overline{z}_k$ are as in (5.45). We use the expression of the coefficients $g_k = \mathbb{E}[W^k]/\mathbb{E}[W^2]^{k/2}$ presented in Lemma 11 of Appendix B.2. When $\mu_W = 0$, then $\lambda = 0$ in (B.1), and the coefficients g_k simplify to $g_k = (k-1)!!\mathcal{E}(k)$, where we recall that $\mathcal{E}(k)$ is the indicator function equal to 1 if k is even, 0 otherwise. If we furthermore define

$$f_k(c,\alpha) := \frac{\alpha/(k-\alpha)}{(\alpha/(2-\alpha))^{k/2}} c^{1-k/2},$$

then the coefficients $\overline{z}_k$ in (5.59) can be accordingly re-written as

$$\overline{z}_k = \frac{i^k}{k!}(k-1)!!\mathcal{E}(k) f_k(c,\alpha), \qquad k \geq 3.$$

Furthermore, using the change variables $p := k/2 \geq 2$ for even $k > 3$, we have[10]

$$\begin{aligned}
\overline{z}_k &= \frac{(-1)^p}{(2p)!}(2p-1)!! f_{2p}(c,\alpha) \\
&= \frac{(-1)^p}{(2p)!!}\frac{c\alpha}{(2p-\alpha)}\left(\frac{c\alpha}{2-\alpha}\right)^{-p} \\
&= \frac{(-1)^p}{p!}\frac{c\alpha}{(2p-\alpha)}\left(\frac{2-\alpha}{2c\alpha}\right)^{p} \\
&= \frac{(-1)^p}{p!}\frac{ca}{(p-a)}\left(\frac{1-a}{2ca}\right)^{p},
\end{aligned}$$

where the identity $(2p)!! = 2^p p!$ is used in the second-last equality, and $a = \alpha/2$ as in (5.57). Then

$$\begin{aligned}
\xi_{c,\alpha}(s) &= ca\sum_{p=2}^{\infty}\frac{(-1)^p}{(p-a)p!}\left(\frac{(1-a)s^2}{2ca}\right)^p \\
&= ca\sum_{p=2}^{\infty}\frac{(-w)^p}{(p-a)p!} \\
&= ca\left(-\frac{e^{-w}}{a} - \frac{w^a\gamma(1-a,w)}{a} + \frac{w}{1-a} + \frac{1}{a}\right) =: \chi_{c,\alpha}(w),
\end{aligned} \tag{5.60}$$

where $w = \frac{(1-a)s^2}{2ca} = \frac{\eta s^2}{2c}$ as in (5.57), $\gamma(s,x)$ is the lower incomplete gamma function (B.10). The last equality can be proved combining the series expansion of $\gamma(s,x)$

$$\gamma(s,x) = x^s\sum_{p=0}^{\infty}\frac{(-x)^p}{(p+s)p!}, \tag{5.61}$$

[10]Notice that $\mathcal{E}(2p) = 1, \forall p \geq 2$.

and the recurrence relation

$$\gamma(s+1,x) = s\gamma(s,x) - x^s e^{-x}, \tag{5.62}$$

for which we refer to Olver et al. [2018]. In fact, (5.60) can be obtained by computing the series expansion (5.61) of $\gamma(-a,w)$, substituting such series in the right-hand side of (5.62), and subtracting the first two terms of the series.[11]
Finally, the statement of the Lemma follows by substituting (5.60) in (5.44) and observing that, using the change of variables (5.57), the CF of the standard normal RV Z (5.27) can be written,

$$\phi_Z(s) = \exp(-s^2/2) = \psi_Z(w) = \exp(-cw/\eta). \tag{5.63}$$

■

Proof: Computing the integral. We start from the expression for $\phi_{R_{(c,d)}}(s)$ in (5.53), where we take $\mu - W = 0$. Performing the change of variables

$$v = \sigma_W^2 s^2 t^2/2, \tag{5.64a}$$
$$t = \sqrt{2v}/\sigma_W|s|, \tag{5.64b}$$
$$dv/dt = \sigma_W^2 s^2 t = \sqrt{2v}\sigma_W|s|, \tag{5.64c}$$

the indefinite integral corresponding to the definite integral in (5.53) with $\mu_W = 0$ may be expanded as

$$\begin{aligned}\int e^{(-\sigma_W^2 s^2 t^2/2)} t^{-\alpha-1}\,\mathrm{d}t &= (|s|\sigma_W)^\alpha 2^{-(\alpha+2)/2}\int \exp(-v)v^{-\alpha/2-1}\,\mathrm{d}v \\ &= (|s|\sigma_W)^\alpha 2^{-(\alpha+2)/2}\left[-\frac{1}{a}v^{-a}e^{-v} - \int \frac{1}{a}v^{-a}e^{-v}\,\mathrm{d}v\right] \\ &= (|s|\sigma_W)^\alpha 2^{-(\alpha+2)/\alpha}\frac{1}{a}\left(-\exp(-v)v^{-a} - \gamma(1-a,v)\right) \\ &= \frac{t^{-\alpha}}{\alpha}\left(-\exp(-v) - v^a\gamma(1-a,v)\right),\end{aligned} \tag{5.65}$$

where $\gamma(s,x)$ is the lower incomplete gamma function (B.10) and $a = \alpha/2$, as in (5.57).
We evaluate the indefinite integral (5.65) in the upper limits

$$t = c^{-1/\alpha}, \qquad v = v_c := \sigma_W^2 s^2 c^{-2/\alpha}/2 \tag{5.66}$$

[11] Observe that this is a variation of the proof of equation (5.60) in Riabiz et al. [2017a], which was based on an differential equation approach.

and in the lower limits

$$t = d^{-1/\alpha}, \qquad v = v_d := \sigma_W^2 s^2 d^{-2/\alpha}/2,$$

respectively, as in (5.54). The definite integral becomes

$$\frac{c}{\alpha}\left(-\exp(-v_c) - v_c^a \gamma(1-a, v_c)\right) - \frac{d}{\alpha}\left(-\exp(-v_d) - v_d^a \gamma(1-a, v_d)\right).$$

Then, as $d \to \infty$, $v_d \to 0$ and $v_d^a \gamma(1-a, v_d) \to 0$, hence the definite integral tends to

$$\frac{c}{\alpha}\left(-\exp(-v_c) - v_c^a \gamma(1-a, v_c)\right) + \frac{d}{\alpha},$$

and substituting in (5.53), recalling that $B = 0$ when $\mu_W = 0$, we have

$$\begin{aligned} \phi_{R_{(c,\infty)}}(s) &= \omega_{R_{(c,\infty)}}(v_c) \\ &= \exp(c(1 - \exp(-v_c) - v_c^a \gamma(1-a, v_c))). \end{aligned} \tag{5.67}$$

Notice that v_c corresponds to u in (5.57), in view of (5.18) and the fact that $\mathbb{E}[W_1^2] = \sigma_W^2$. Then the statement follows by recalling that, from (5.42),

$$\phi_{Z_{(c,\infty)}}(s) = \phi_{R_{(c,\infty)}}(s/S_{(c,\infty)}).$$

Finally, using the change of variables (5.57), $\phi_{R_{(c,\infty)}}(s/S_{(c,\infty)}) = \omega_{R_{(c,\infty)}}(u/S_{(c,\infty)}^2) = \psi_{R_{(c,\infty)}}(w)$. ■

Analytic Expression for $\phi_{X_{(0,c)}}(s)$

Using computations similar to those in the second proof of Lemma 5, it is also possible to derive closed-form expressions for the CF of the truncated PSR $X_{(0,c)}$, as follows.

Lemma 6. Suppose $W_1 \sim \mathcal{N}(0, \sigma_W^2)$, and let a and u be defined as in (5.57), $\alpha \in (0,2), \alpha \neq 1$. Then,

$$\begin{aligned} \phi_{X_{(0,c)}}(s) &= \omega_{X_{(0,c)}}(u) \\ &= \exp(-c(1 - e^{-u} + u^a \Gamma(1-a, u))), \end{aligned} \tag{5.68}$$

where $\Gamma(s,x)$ is the upper incomplete gamma function (B.11).

Proof: To prove the Lemma, we evaluate the indefinite integral (5.65) in the upper limits $t = \infty$ and $v = \infty$, and in the lower limits now equal to (5.66)

(the upper limits for the CF of $R_{(c,\infty)}$) as in (5.55). Recalling from (5.64a) that $t^{-\alpha}v^a = (|s|\sigma_W/\sqrt{2})^\alpha$, where $\alpha = 2a$, as in (5.57), and also recalling the definition of the gamma function (B.12), the corresponding definite integral becomes

$$-\frac{1}{\alpha}\left(\frac{|s|\sigma_W}{\sqrt{2}}\right)^\alpha \Gamma(1-a) - \frac{c}{\alpha}\left(-\exp(-v_c) - v_c^a\gamma(1-a, v_c)\right),$$

and, substituting in (5.55) and using the definition of v_c,[12] we obtain

$$\begin{aligned}\phi_{X_{(0,c)}}(s) &= \omega_{X_{(0,c)}}(v_c)\\ &= \exp(-c(1-\exp(-v_c) - u_c^a\gamma(1-a, v_c) + v_c^a\Gamma(1-a)))\\ &= \exp(-c(1-\exp(-v_c) + v_c^a\Gamma(1-a, v_c))),\end{aligned}$$

as claimed. ■

Analytic Expressions for $\phi_X(s)$ and $\phi_{\hat{X}}(s)$

Here we check that, if the unnormalized residual $R_{(c,\infty)}$ is added to the truncated series $X_{(0,c)}$, we get the regular symmetric stable CF, as expected. By multiplying the CFs (5.67) and (5.68) together, as in (5.40), we obtain

$$\begin{aligned}\phi_X(s) &= \phi_{X_{(0,c)}}(s)\phi_{R_{(c,\infty)}}(s)\\ &= \exp(-c(1-\exp(-v_c) + v_c^a\Gamma(1-a, v_c)) + c(1-\exp(-v_c) - v_c^a\gamma(1-a, v_c)))\\ &= \exp(-cv_c^a\Gamma(1-a))\\ &= \exp(-\sigma_W^\alpha 2^{-a}\Gamma(1-a)|s|^\alpha),\end{aligned} \tag{5.69}$$

with a as in (5.57), v_c as in (5.66), and $\Gamma(1-a)$ the gamma function (B.12). Then, using equation (A.2) from Appendix A.1, we have that $-\sigma_W^\alpha 2^{-a}\Gamma(1-a) = -\sigma^\alpha$ making (5.69) coincide with the stable CF (2.3) in the symmetric case ($\beta = 0$).

Furthermore, from (5.63) and (5.43), it follows that, when $\mu_W = 0$

$$\phi_{\hat{R}_{(c,\infty)}}(s) = \omega_{\hat{R}_{(c,\infty)}}(u) = \exp(-cu/\eta). \tag{5.70}$$

Then, as a consequence of Lemma 6 and equation (5.41), we have that, when $\mu_W = 0$, the CF of $\hat{X}$, the approximated stable distribution is $\phi_{\hat{X}}(s) = \omega_{\hat{X}}(u)$, such that

$$\log\omega_{\hat{X}}(u) = -c(1-e^{-u} + u^a\Gamma(1-a, u) + u/\eta). \tag{5.71}$$

[12] Observe that v_c corresponds to u in (5.57).

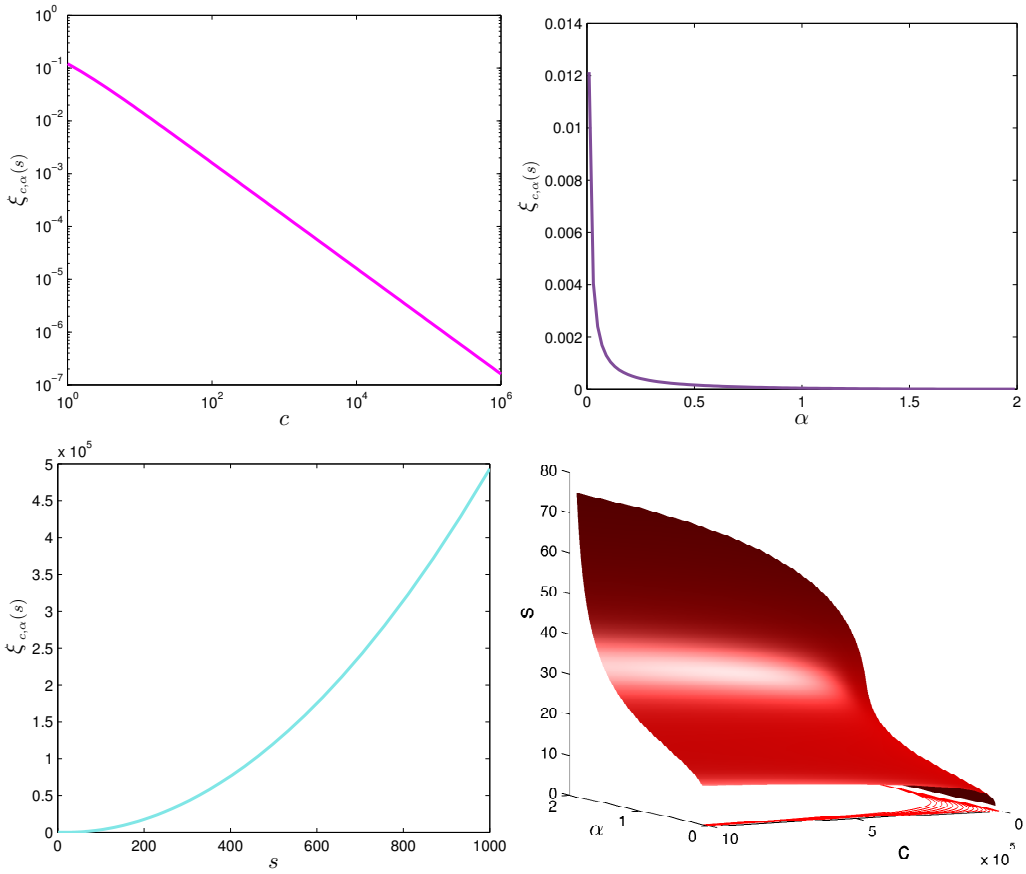


Figure 5.5 Top-left: $\xi_{c,\alpha}(s)$ as a function of c. Top-right: $\xi_{c,\alpha}(s)$ as a function of α. Bottom-left: $\xi_{c,\alpha}(s)$ as a function of s. Bottom-right: contour-surface for a fixed value of $\xi_{c,\alpha}(s) = 0.01$. In all subfigures, when not stated $s = 1$, $\alpha = 0.5$ and $c = 1000$.

Figure 5.5 represents the function $\xi_{c,\alpha}(s)$ (5.60), which constitutes the 'extra-Gaussian' part of the CF of the PSR residual when $W_j \sim \mathcal{N}(0, \sigma_W^2)$, see (5.44) and (5.59). The observed decay of $\xi_{c,\alpha}(s)$ suggests that the distance between the CDF of the standardized PSR residual and the Gaussian CDF also decreases, as c is increased and $\alpha \to 2$.[13] Transferring the information on distance between CFs to the distance of CDFs is the aim of the following chapters.

[13] This is also in line with the Monte Carlo simulations presented in Section 5.3.1.

Chapter 6

Nonasymptotic Gaussian Bounds for the PSR Residual

In this Chapter, we provide analytic nonasymptotic bounds on the Kolmogorov distance $\Delta(Z_{(c,\infty)}, Z)$ (5.36), as a function of the truncation parameter c and the tail parameter α. We restrict our attention to the case $\mu_W = 0$, corresponding to the symmetric stable law, for which we have the closed-form expressions for $\phi_{Z_{(c,\infty)}}(s)$ derived in the previous Chapter. We leave to future work the extension of these theoretical results to a generally skewed stable law, which is the ultimate target when using the PSR for obtaining a MaSMiN model (recall that the simpler SMiN model could be applied directly for the symmetric case).

The Chapter is structured as follows: we first provide some preliminary results that are used in the Chapter, including the study of the logarithm of the CF of $Z_{(c,\infty)}$. We then present a bound on $\Delta(Z_{(c,\infty)}, Z)$ whose asymptotic order is $O(1/c)$, derived by resorting to the smoothing lemma with the smoothing parameter set to infinity (5.35). Such bound states an asymptotically fast decay, but it is lose for small values of c and α. Thus a second bound is obtained by using the smoothing lemma with finite smoothing parameter (5.34): such bound has asymptotic order $O(1/\sqrt{c})$, but that is sharper than the first bound for small c and α. The combination of these two forms our overall analytic bound on $\Delta(Z_{(c,\infty)}, Z)$, that it is compared to numerical simulations. We conclude the Chapter reporting numerical simulations on the distance between the PDF of the PSR residual and the Gaussian PDF, according to the continuity theorem for densities.

The results have been developed in collaboration with Prof. Ioannis Kontoyiannis, and they were presented in Riabiz et al. [2017c, 2018b] without proof and then in Riabiz et al. [2018a] with proof, respectively.

6.1 Preliminary Results

In this Section we give some preliminary results that are used in the rest of this Chapter, as well as in Chapter 7.

6.1.1 Difference of Powers

The following lemma is a simple calculus exercise.

Lemma 7. If $0 \leq x \leq C$ and $0 \leq y \leq C$, then for any (not necessarily integer) $n \geq 1$

$$|x^n - y^n| \leq n|x - y|C^{n-1}.$$

Proof: Without loss of generality, assume $0 \leq y \leq x \leq C$, and fix x. Then the statement is equivalent to proving $\zeta(y) := x^n - y^n - n(x-y)C^{n-1} < 0$. This is easy to check, because $\zeta(0) = x^n - nxC^{n-1} = x(x^{n-1} - nC^{n-1}) \leq xC^{n-1}(1-n) \leq 0$, and $\zeta'(y) = nC^{n-1} - ny^{n-1} \geq 0$, while $\lim_{y\to x} \zeta(y) = 0$. ■

6.1.2 Study of the Function $g(w)$

When $W_1 \sim \mathcal{N}(0, \sigma_W^2)$, (5.58) provides the expression of the full CF of the PSR standardized residual. For convenience we write

$$\psi_{Z_{(c,\infty)}}(w) = f(w)^c, \tag{6.1}$$

with

$$f(w) := \exp\left(1 - e^{-w} - w^a \gamma\left(1 - a, w\right)\right). \tag{6.2}$$

Then, in this section, we study the function

$$g(w) := \log f(w) = 1 - e^{-w} - w^a \gamma(1-a, w). \tag{6.3}$$

Upper Bounds on $g(w)$

First note that $g(0) = 0$ and that, from Lemma 16, we have $g(w) \leq 0$ for all $w \geq 0$ and in fact,

$$g(w) \leq -\frac{1 - e^{-w}}{\eta} \leq -\frac{w}{\eta}\left(1 - \frac{w}{2}\right), \tag{6.4}$$

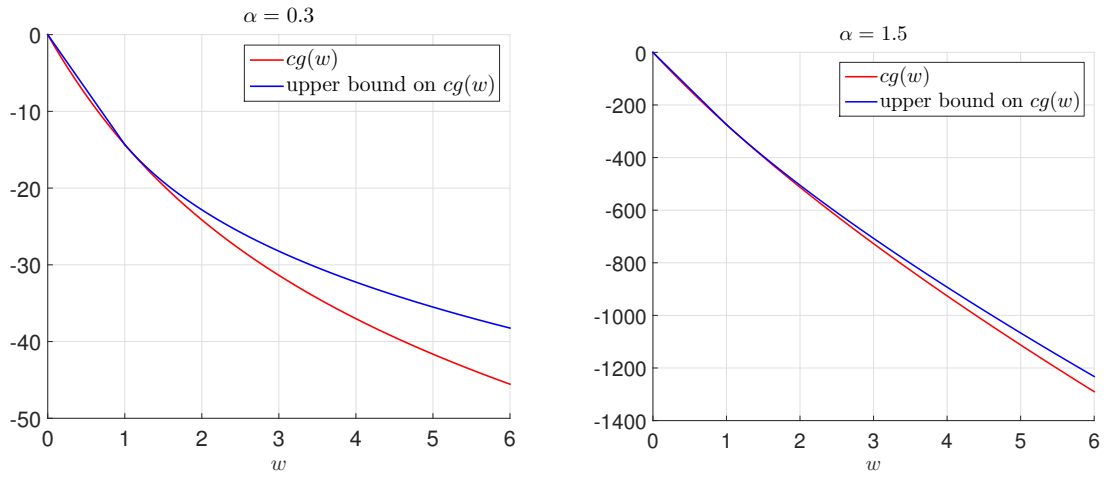


Figure 6.1 The red lines are examples of the function $cg(w)$), with $c = 100$ and $\alpha \in \{0.3, 1.5\}$). The blue lines represent the upper bound $\bar{g}w\mathbb{1}(w \in (0,1)) + (1 - e^{-1} - \bar{\gamma}(a)w^a)\mathbb{1}(w > 1)$.

where the second step follows from the fact that $e^{-x} \leq 1 - x + x^2/2$, for $x \geq 0$. Therefore,

$$g(w) \leq -\frac{w}{2\eta}, \quad 0 \leq w \leq 1. \tag{6.5}$$

On the other hand, for $w \geq 1$ we have the bound:

Lemma 8. For all $w \geq 1$:

$$g(w) \leq 1 - e^{-1} - \bar{\gamma}(a)w^a, \tag{6.6}$$

where

$$\bar{\gamma}(a) := \gamma(1 - a, 1). \tag{6.7}$$

Proof: The statement is equivalent to,

$$H(w) := \gamma(1 - a, w) - \gamma(1 - a, 1) - \frac{e^{-1} - e^{-w}}{w^a} \geq 0,$$

for all $w \geq 1$. Since $H(1) = 0$ and, using (B.18), the derivative

$$H'(w) = aw^{-a-1}(e^{-1} - e^{-w}),$$

is always nonnegative, the result follows. ■

Monotonicity and Lower Bound

Differentiating,

$$g'(w) = -aw^{a-1}\gamma(1-a,w), \tag{6.8}$$

so that $g'(0+) = -1/\eta$ by (B.14), and,

$$0 \geq g'(w) \geq -\frac{1+(1-a)e^{-w}}{\eta(2-a)} \geq -\frac{1}{\eta}, \tag{6.9}$$

for all $w \geq 0$, where the first inequality is obvious by (6.8), the second follows from Lemma 15, and the third from the fact that $e^{-w} \leq 1$ always. In particular, this implies that,

$$g(w) \geq -\frac{w}{\eta}, \quad w \geq 0. \tag{6.10}$$

Concavity and Further Upper Bounds

Differentiating again,

$$g''(w) = -\frac{a}{w}\left[e^{-w} - (1-a)w^{a-1}\gamma(1-a,w)\right],$$

and applying Lemma 17, we have that,

$$-a \leq \frac{a(1-w)}{2} \leq g''(w) \leq \frac{a}{2-a}, \tag{6.11}$$

where the first inequality only holds for $0 \leq w \leq 3$; on the other hand, since (B.15) holds for all $w \geq 0$ and the function $x \mapsto (x+1)e^{-x}$ is decreasing, for all $w \geq 1$ we have the simple lower bound,

$$g''(w) \geq -\frac{a}{w}\left(\frac{(w+1)e^{-w}-1}{w}\right) \geq \frac{a(1-2e^{-1})}{w^2}. \tag{6.12}$$

In particular, from (6.11) and (6.12) it follows that,

$$0 \leq g''(w) \leq \frac{a}{2-a}, \quad \text{for all } w \geq 0. \tag{6.13}$$

This implies that the function $g(w)$ is *convex*, and a tighter upper bound than (6.5) for $w \in [0,1]$ is the linear interpolation at the extremes:

$$g(w) \leq \bar{g}w, \quad 0 \leq w \leq 1, \tag{6.14}$$

where

$$\bar{g} := g(1), \tag{6.15}$$

More generally, the linear interpolating upper bound for $g(w)$ on $[0, \bar{w}]$ is

$$g(w) \leq \frac{g(\bar{w})}{\bar{w}} w, \quad 0 \leq w \leq \bar{w}. \tag{6.16}$$

Figure 7.1 shows two examples of the function $cg(w) = \log(\psi_{Z_{(c,\infty)}}(w))$ and of its bound $\bar{g}w\mathbb{1}(w \in (0,1)) + (1 - e^{-1} - \bar{\gamma}(a)w^a)\mathbb{1}(w > 1)$, for two different values of α.

6.2 Nonasymptotic Bound of Order $O(1/c)$

In this Section we apply the smoothing lemma with with the smoothing parameter set to infinity (5.35), in order to derive explicit bounds on the distance between the CDF of the standardized PSR residual and the standard Gaussian CDF. In particular, we prove the following:

Theorem 6.2.1. Let $W_j \sim \mathcal{N}(0, \sigma_W^2)$ and let $\Delta(Z_{(c,\infty)}, Z)$ be the Kolmogorov distance between $Z_{(c,\infty)}$ and Z, as in (5.36). Let $a = a(\alpha)$ and $\eta = \eta(\alpha)$ as in (5.57), and $\bar{\gamma}(a)$ as in (6.7), $\bar{g}$ as in (6.15), and[1]

$$K(a) := \frac{1}{\pi}\left[\frac{a}{2(2-a)} + \frac{1}{\eta^2}\right]. \tag{6.17}$$

Then, for any $c > 1$, $\Delta(Z_{(c,\infty)}, Z)$ is bounded above by:

$$\begin{aligned} B_1(c,\alpha) := \frac{cK(a)}{(c-1)}\Bigg[\frac{1}{(c-1)\bar{g}^2} + \left(\frac{1}{\bar{g}} - \frac{1}{(c-1)\bar{g}^2}\right)\exp\left((c-1)\bar{g}\right) + \\ + \frac{(c-1)\exp\{(c-1)(1-e^{-1})\}}{a[(c-1)\bar{\gamma}(a)]^{2/a}}\Gamma\left(2/a, (c-1)\bar{\gamma}(a)\right)\Bigg], \end{aligned} \tag{6.18}$$

where $\gamma(s,x)$ and $\Gamma(s,x)$ are the lower and upper incomplete gamma functions, (B.10) and (B.11) respectively.

Proof: For convenience we divide the proof in the following four steps.
Step I. Recall that, when $W_j \sim \mathcal{N}(0, \sigma_W^2)$, the CF of $Z_{(c,\infty)}$ can be expressed as $\psi_{Z_{(c,\infty)}}(w) = f(w)^c$, see (6.1), with $f(w)$ defined as in (6.2). We apply the smoothing Lemma (5.34) to $\Delta(Z_{(c,\infty)}, Z)$ as in (5.36). Given that the standard Gaussian PDF

[1] Observe that, with an abuse of notation, here we mean a function different from (2.5).

is uniformly bounded by $m := 1/\sqrt{2\pi} < 2/5$, for any $\Theta > 0$ equation (5.34) gives

$$\pi\Delta(Z_{(c,\infty)}, Z) \leq \int_{-\Theta}^{\Theta} |\phi_{Z_{(c,\infty)}}(s) - \phi_Z(s)| \frac{1}{|s|}\, \mathrm{d}s + \frac{9.6}{\Theta}, \tag{6.19}$$

with $\phi_Z(s)$ as in (5.27). Letting $\Theta \to \infty$ and changing variables according to (5.57)

$$\begin{aligned} \pi\Delta(Z_{(c,\infty)}, Z) &\leq \int_0^{\infty} \left| f(w)^c - \phi_Z\Big(\sqrt{2w/\eta}\Big)^c \right| \frac{1}{w}\, \mathrm{d}w \\ &=: \pi \bar{I}(Z_{(c,\infty)}, Z). \end{aligned} \tag{6.20}$$

Step II. We apply Lemma 7 to the integrand in (6.20), with $x = f(w)$, $y = \phi_Z(\sqrt{2w/\eta}) = \exp(-w/\eta)$ by (5.63), and $n = c$. Let $g(w) = \log(f(w))$ as in (6.3), and $\bar{\gamma}(a)$ and $\bar{g}$ as in (6.7) and (6.15), respectively, and define

$$h(w) := \begin{cases} -\bar{g}w, & w \in [0,1] \\ e^{-1} - 1 + \bar{\gamma}(a) w^a, & w > 1. \end{cases} \tag{6.21}$$

From (6.14) and (6.6) it follows that $x \leq \exp(-h(w))$ for all $w \geq 0$. Since by (6.10) $y \leq x$, $\forall w \geq 0$, it follows immediately $y \leq \exp(-h(w))$, $\forall w \geq 0$. Therefore, we can take $C = \exp(-h(w))$ in Lemma 7, and substituting the resulting bound in (6.20) we get

$$\pi\Delta(Z_{(c,\infty)}, Z) \leq \int_0^{\infty} \frac{c}{w} \left| f(w) - \phi_Z\left(\sqrt{\frac{2w}{\eta}}\right) \right| e^{-(c-1)h(w)}\, \mathrm{d}w. \tag{6.22}$$

Step III. In order to bound the absolute difference in the integrand in (6.22), we express f as a quadratic Taylor expansion. Noting that $f' = g' \exp(g)$ and $f'' = (g'' + g'^2)\exp(g)$, where g is defined in (6.3), and recalling that $g(0) = 0$ and $g'(0) = -1/\eta$, we have

$$\begin{aligned} \left| f(w) - 1 + \frac{w}{\eta} \right| &\leq \frac{w^2}{2} \sup_{v \geq 0} \left[\Big(|g''(v)| + g'(v)^2 \Big)\, e^{g(v)} \right] \\ &\leq \frac{w^2}{2} \left[\frac{a}{2-a} + \frac{1}{\eta^2} \right], \end{aligned}$$

where in the second inequality we used the earlier bounds in (6.9) and (6.13), and also the fact that $g(w) \leq 0$ for all $w \geq 0$, see (6.4). From the standard quadratic

expansion for the exponential function we similarly have

$$\left|\exp\left(-\frac{w}{\eta}\right) - 1 + \frac{w}{\eta}\right| \leq \frac{w^2}{2\eta^2}.$$

Combining the last two bounds,

$$\left|f(w) - \phi_Z\left(\sqrt{\frac{2w}{\eta}}\right)\right| \leq \frac{w^2}{2}\left[\frac{a}{2-a} + \frac{2}{\eta^2}\right]. \tag{6.23}$$

Step IV. Finally, substituting (6.23) in (6.22),

$$\begin{aligned}\Delta(Z_{(c,\infty)}, Z) &\leq \frac{c}{\pi}\left[\frac{a}{2(2-a)} + \frac{1}{\eta^2}\right]\int_0^\infty w\exp\Big(-(c-1)h(w)\Big)\,\mathrm{d}w \\ &= cK(a)[I^Z(c) + J^Z(c)],\end{aligned} \tag{6.24}$$

where $K(c)$ is defined in (6.17), and where we first integrate over $[0,1]$,

$$\begin{aligned}I^Z(c) &:= \int_0^1 w\exp\Big(w(c-1)\bar{g}\Big)\,\mathrm{d}w \\ &= \frac{1}{(c-1)^2\bar{g}^2} + \frac{1}{(c-1)\bar{g}}\left(1 - \frac{1}{(c-1)\bar{g}}\right)\exp\left((c-1)\bar{g}\right),\end{aligned} \tag{6.25}$$

and then over $[1,\infty)$,

$$\begin{aligned}J^Z(c) &:= e^{(1-e^{-1})(c-1)}\int_1^\infty w\exp\Big\{-(c-1)\bar{\gamma}(a)w^a\Big\}\,\mathrm{d}w \\ &= \frac{e^{(1-e^{-1})(c-1)}}{a[(c-1)\bar{\gamma}(a)]^{2/a}}\int_{(c-1)\bar{\gamma}(a)}^\infty u^{\frac{2}{a}-1}e^{-u}\,\mathrm{d}u \\ &= e^{(1-e^{-1})(c-1)}\frac{\Gamma(2/a, (c-1)\bar{\gamma}(a))}{a[(c-1)\bar{\gamma}(a)]^{2/a}},\end{aligned} \tag{6.26}$$

where $\Gamma(s,x)$ is the upper incomplete gamma function defined in (B.11). Combining (6.24), (6.25) and (6.26) gives the claimed result, after re-writing

$$\begin{aligned}B_1(c,\alpha) &= cK(a)\Big(I^Z(c) + J^Z(c)\Big) \\ &= \frac{cK(a)}{c-1}\Big((c-1)I^Z(c) + (c-1)J^Z(c)\Big)\end{aligned} \tag{6.27}$$

for convenience. ■

Remark 7. Observe that the upper bound provided by the smoothing lemma on the Kolmogorov distance of the PSR residual from its Gaussian approximation,

$\bar{I}(R_{(c,\infty)}, \hat{R}_{(c,\infty)})$, coincides with the upper bound on the the distance of the PSR standardized residual from the standard Gaussian, $\bar{I}(Z_{(c,\infty)}, Z)$ as in (6.20). In fact

$$\begin{aligned}\bar{I}(R_{(c,\infty)}, \hat{R}_{(c,\infty)}) &= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\left|\phi_{R_{(c,\infty)}}(s) - \phi_{\hat{R}_{(c,\infty)}}(s)\right|}{|s|}\,\mathrm{d}s \\ &= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\left|\phi_{Z_{(c,\infty)}}(sS_{(c,\infty)}) - \phi_Z(sS_{(c,\infty)})\right|}{|s|}\,\mathrm{d}s \\ &= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\left|\phi_{Z_{(c,\infty)}}(s) - \phi_Z(s)\right|}{|s|}\,\mathrm{d}s \\ &= \bar{I}(Z_{(c,\infty)}, Z),\end{aligned}$$

where in the second identity we have used (5.42) and (5.43) and the fact that $\left|\exp(ism_{(c,\infty)})\right| = 1$, while the third identity trivially follows through a change of variable.

Remark 8. In general, when $W_1 \sim \mathcal{N}(\mu_W, \sigma_W^2)$, the CF of $Z_{(c,\infty)}$ will depend on μ_W and σ_W^2, or rather on $\lambda = \mu_W/\sigma_W$. This is clear considering the expression (5.44) of $\phi_{Z_{(c,\infty)}}(s)$ and that of the coefficients g_k in (B.1). If $\mu_W = 0$, as in our case, then $\lambda = 0$, making these bounds hold the same for any value of σ_W, namely for the PSR residual of any symmetric α-stable distribution, regardless of σ.

Remark 9. Even though the PSR split that we introduced in (5.1) is defined for any $c \geq 0$, Theorem 6.2.1, as well as the following results, hold on for $c > 1$.[2] We do not consider this to be a significant limitation because, in practice, we are indeed interested in scenarios where there are at least a few terms in $X_{(0,c)}$, compatible with $c > 1$ (recall that $\mathbb{E}[N_{(0,c)}] = c$).

The following corollary is a simple consequence of Theorem 6.2.1;

Corollary 6.2.2. Under the same assumptions of Theorem 6.2.1, as $c \to \infty$,

$$B_1(c, \alpha) \sim \frac{K(a)}{\bar{g}^2}\left(\frac{1}{c-1}\right),$$

so that

$$\Delta(Z_{(c,\infty)}, Z) = O(1/c)$$

.

[2] In fact the expression of $B_1(c, \alpha)$ requires $c > 1$, even if it is derived applying Lemma 7, that allows $c \geq 1$.

Proof: It is clearly sufficient to prove the first asymptotic assertion of the corollary. For this purpose, we take the limit of each of the terms in th expression (6.27) for $B_1(c, \alpha)$. We have

$$\frac{c}{c-1}K(a) \sim K(a),$$

and

$$\begin{aligned}(c-1)I^Z(c) &= \frac{1}{(c-1)\bar{g}^2} + \left(\frac{1}{\bar{g}} - \frac{1}{(c-1)\bar{g}^2}\right)\exp\left((c-1)\bar{g}\right) \\ &\sim \frac{1}{(c-1)\bar{g}^2},\end{aligned}$$

given that $\bar{g} < 0$. Finally

$$\begin{aligned}(c-1)J^Z(c) &= \frac{(c-1)\exp\left\{(c-1)(1-e^{-1})\right\}}{a[(c-1)\bar{\gamma}(a)]^{2/a}}\Gamma\left(2/a, (c-1)\bar{\gamma}(a)\right) \\ &\sim \frac{(c-1)\exp\left\{(c-1)(1-e^{-1})\right\}}{a[(c-1)\bar{\gamma}(a)]^{2/a}} \times [(c-1)\bar{\gamma}(a)]^{2/a-1}\exp(-(c-1)\bar{\gamma}(a)) \\ &= \frac{1}{a\bar{\gamma}(a)}\exp\left\{(c-1)(1-e^{-1}-\bar{\gamma}(a))\right\},\end{aligned}$$

where, in solving the asymptotic we have used the property (B.17). Observe that the exponent in the last expression is negative. In fact, applying (B.13) to $\bar{\gamma}(a)$ we have

$$1 - e^{-1} - \bar{\gamma}(a) < (1-e^{-1})\left(1 - \frac{1}{1-a}\right) < 0,$$

given that $a \in (0, 1)$, as in (5.57). Therefore, combining these estimates with (6.27), we indeed have

$$B_1(c, \alpha) \sim K(a)\frac{1}{(c-1)g^2(1)}.$$

■

An important remark is that the asymptotic order of the bound on $\Delta(Z_{(c,\infty)}, Z)$ is independent of the tail parameter α, which we will show to be consistent with numerical simulations of the smoothing lemma bound $\bar{I}(Z_{(c,\infty)}, Z)$ (6.20)

For values of α greater than 0.4, $B_1(c, \alpha)$ gives very good bounds, as shown on the left-hand side of Figure 6.2. But for α below 0.4, the results deteriorate significantly; for example, for $\alpha = 0.2$, $B_1(c, \alpha)$ is below 1 (the maximum possible Kolmogorov distance) only for $c > 115$.

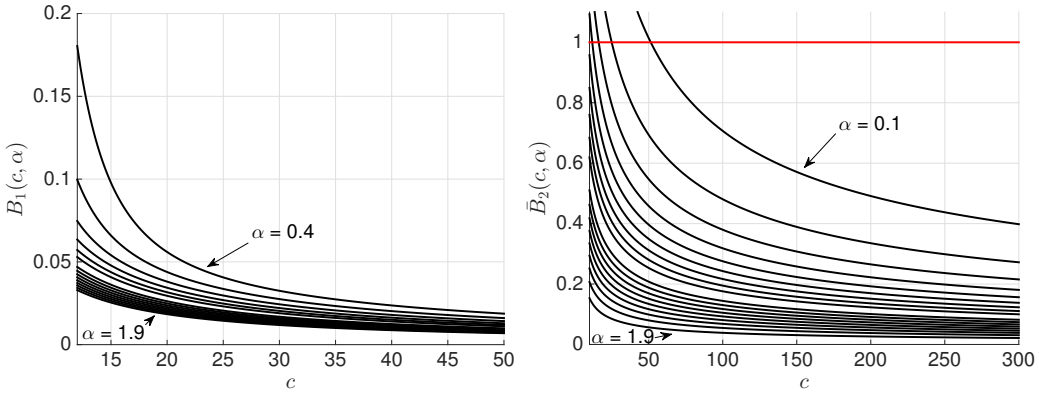


Figure 6.2 Bounds on $\Delta(Z_{(c,\infty)}, Z)$. Left: each curve represents the values of the bound $B_1(c,\alpha)$, of order $O(1/c)$, for $\alpha = 0.4, 0.9, \ldots, 1.9$, $\alpha \neq 1$, plotted against $12 \leq c \leq 50$. Right: each curve represents the values of the bound $\bar{B}_2(c,\alpha)$, of order $O(1/\sqrt{c})$, for $\alpha = 0.1, 0.2, \ldots, 1.9$, $\alpha \neq 1$, plotted against $10 \leq c \leq 300$. The red horizontal line is simply equal to 1, the maximum possible value of the Kolmogorov distance.

6.3 Nonasymptotic Bound of Order $O(1/\sqrt{c})$

The following result, obtained by bounding $I(Z_{(c,\infty)}, Z)$ in (5.34), gives an $O(1/\sqrt{c})$ bound which is, of course, asymptotically inferior to that in Theorem 6.2.1, but which gives sharper results for small c and $\alpha < 0.4$.

Theorem 6.3.1. Under the same assumptions and in the notation of Theorem 6.2.1, for any $\delta \in (0,2)$, $\Delta(Z_{(c,\infty)}, Z)$ is bounded above by,

$$B_2(c,\alpha,\delta) := \frac{9.6\sqrt{\eta}}{\pi\sqrt{2(2-\delta)c}} + B_3(c,\alpha,\delta),$$

where $B_3(c,\alpha,\delta)$ is the following $O(1/c)$ term:

$$B_3(c,\alpha,\delta) := \frac{K(a)}{c}\left(\frac{c(2-\delta)}{(c-1)g(2-\delta)}\right)^2 \times \left\{1 - \left[1 - g(2-\delta)(c-1)\right]\exp\left(g(2-\delta)(c-1)\right)\right\}.$$

Proof: As in the proof of Theorem 6.2.1, we start from (6.19) and we perform the change of variables (5.57) to obtain that, for any $\Theta > 0$,

$$\pi\Delta(Z_{(c,\infty)}, Z) \leq \int_0^{\eta\Theta^2/2c} \left| f(w)^c - \phi_Z\left(\sqrt{\frac{2w}{\eta}}\right)^c \right| \frac{1}{w}\,\mathrm{d}w + \frac{9.6}{\Theta}. \tag{6.28}$$

Instead of letting $\Theta \to \infty$, we choose $\Theta = \sqrt{2(2-\delta)c/\eta}$, and we apply Lemma 7 to the integrand in (6.28), with $x = f(w)$ and $y = \phi_Z(\sqrt{2w/\eta}) = \exp(-w/\eta)$ via (5.57). From (6.10) $y \leq x$, for all $w \geq 0$, and using (6.16) with $\bar{w} = 2 - \delta$ we have

that $x \leq \exp\left(\frac{g(2-\delta)}{2-\delta}w\right)$. We can then take $C = \exp\left(\frac{g(2-\delta)}{2-\delta}w\right)$ and Lemma 7 applied to (6.28) gives

$$\pi\Delta(Z_{(c,\infty)}, Z) - \frac{9.6\sqrt{\eta}}{\sqrt{2(2-\delta)c}} \leq \int_0^{2-\delta} \frac{c}{w} \left| f(w) - \phi_Z\left(\sqrt{\frac{2w}{\eta}}\right)\right| \exp\left(\frac{(c-1)g(2-\delta)w}{2-\delta}\right) \mathrm{d}w. \tag{6.29}$$

Recalling the earlier expansion (6.23), substituting in (6.29), and integrating, yields

$$\begin{aligned}\pi\Delta(Z_{(c,\infty)}, Z) - \frac{9.6\sqrt{\eta}}{\sqrt{2(2-\delta)c}} &\leq c\left[\frac{a}{2(2-a)} + \frac{1}{\eta^2}\right] \int_0^{2-\delta} w \exp\left(\frac{(c-1)g(2-\delta)w}{2-\delta}\right) \mathrm{d}w \\ &= \frac{1}{c}\left[\frac{a}{2(2-a)} + \frac{1}{\eta^2}\right] \left(\frac{c(2-\delta)}{(c-1)g(2-\delta)}\right)^2 \times \\ &\quad \times \left\{1 - [1 - g(2-\delta)(c-1)] \exp\left(g(2-\delta)(c-1)\right)\right\},\end{aligned}$$

as claimed.

■

Numerically minimizing the bound $B_2(c, \alpha, \delta)$ over δ yields $\bar{B}_2(c, \alpha)$, shown in the right-hand side of Figure 6.2.

6.4 Combined Bound and Comparison with Numerical Results

Finally, we combine the results of Theorem 6.2.1 and Theorem 6.3.1, to obtain useful bounds essentially for all values of $\alpha \in (0, 2)$, $\alpha \neq 1$, and $c > 50$ as,

$$B_4(c, \alpha) := \min\left\{B_1(c, \alpha), \bar{B}_2(c, \alpha)\right\}.$$

The result is shown in the left-hand side of Figure 6.3, at a log-log scale.

The right-hand side of Figure 6.3 shows a comparison between the theoretical bound $B_4(c, \alpha)$ and the numerical estimate $\bar{Q}(Z_{(c,\infty)}, Z)$ of $\bar{I}(Z_{(c,\infty)}, Z)$ (6.20) that was first presented in Riabiz et al. [2017c] and Riabiz et al. [2018b]. The numerical values are produced through the Matlab routine `quadgk`, which implements the Gauss-Kronrod method. For better numerical stability, we use (5.57)[3] and work with CF expressed wrt the variables w. The Gauss-Kronrod scheme attempts to approximate the integral of a scalar-valued function on a bounded interval using

[3]Consistently with the considerations in Remark 8, it is not necessary to specify the value of σ_W^2 for the numerical simulations.

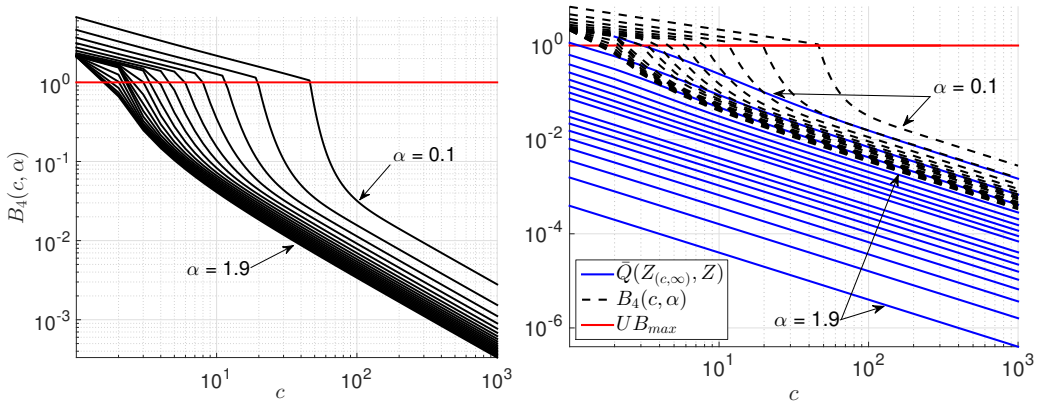


Figure 6.3 Bounds on $\Delta(Z_{(c,\infty)}, Z)$. Left: each curve represents the combined bound $B_4(c,\alpha)$, for $\alpha = 0.1, 0.2, \ldots, 1.9$, $\alpha \neq 1$, plotted against $10 \leq c \leq 1000$. Right: comparison between $\bar{Q}(Z_{(c,\infty)}, Z)$ and $B_4(c,\alpha)$, for $\alpha = 0.1, 0.2, \ldots, 1.9$, $\alpha \neq 1$, plotted for $3 \leq c \leq 300$. The red horizontal lines are simply equal to 1, the maximum possible value of the Kolmogorov distance.

high-order global adaptive quadrature. Infinite intervals (such as in our case) are dealt with using changes of variables; discontinuity points ($s = 0$ or, equivalently, $w = 0$, in our case) need to be passed as input to the routine. The method outputs also an approximate upper bound on the absolute error $|\bar{I}(Z_{(c,\infty)}, Z) - \bar{Q}(Z_{(c,\infty)}, Z)|$, which can be used to construct approximate error bands. The numerical quadrature encounters some difficulties when trying to approximate $\bar{I}(Z_{(c,\infty)}, Z)$ for $c \in (1, 3)$, and $\alpha \leq 0.2$: either the method has not converged, or the error bands are very large. However, the numerical error intervals are negligibly small for $c \geq 3$, so we do not show them here. Observe that $B_4(c,\alpha)$ appears to have the same asymptotic rate as $\bar{Q}(Z_{(c,\infty)}, Z)$. Moreover, the values of $B_4(c,\alpha)$ for a fixed truncation parameter c are smaller as $\alpha \to 2$, as we expected from the Monte Carlo simulations in Chapter 5.

6.5 Numerical Results for the PDF Bounds

We conclude this chapter reporting our numerical simulations for the bounds on the distance between the PDF of $Z_{(c,\infty)}$ and the standard Gaussian PDF for the symmetric stable distribution, with $W_1 \sim \mathcal{N}(0, \sigma_W^2)$ in the PSR. The bounds are obtained simulating the integral $i(Z_{(c,\infty)}, Z)$, defined as in (5.39), through the Matlab routine `quadgk`, which gives the numerical approximation $q(Z_{(c,\infty)}, Z)$, represented in Figure 6.4. In the figure we are also plotting the approximate error bands on $|i(Z_{(c,\infty)}, Z) - q(Z_{(c,\infty)}, Z)|$, but they are negligibly small. A regression on the lines shows that the asymptotic order of $q(Z_{(c,\infty)}, Z)$ is also $O(1/c)$. As observed in Section 5.3.3, the bounds obtained in this way are uniform wrt the variable of the PDFs.

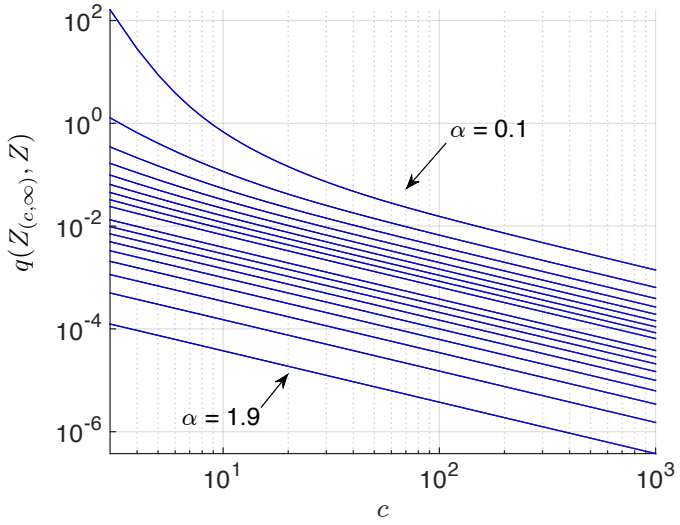


Figure 6.4 Numerical bounds on the PDF divergence $\delta(Z_{(c,\infty)}, Z)$, as a function of $3 \leq c \leq 1000$ and $\alpha = 0.1, \ldots, 1.9$, $\alpha \neq 1$, The plots include approximate error bands for each line.

We leave to future work the derivation of analytic bounds for $\delta(Z_{(c,\infty)}, Z)$, based on (5.39), and, more importantly, the formulation of non-uniform bounds; the latter appears to be promising the development of novel inference schemes based on the PSR, as we elaborate in Chapter 9.

Now that we have obtained bounds on the distance of the PSR residual from a Gaussian, we can proceed on bounding the distance of the PSR approximation from the stable law corresponding to the full PSR.

Chapter 7

Nonasymptotic Bounds on the Approximation of the PSR

In this Chapter, we provide and compare bounds on: (i) the Kolmogorov distance between the stable law, and its approximation stemming from the GAA scheme for the PSR, $\Delta(X, \hat{X})$, defined in (5.37); (ii) the Kolmogorov distance between the stable law, and its approximation consisting of simple truncation of the PSR $\Delta(X, X_{(0,c)})$, defined in (5.38). For both distances, we consider the case of symmetric stable distributions, in which we have closed-form expressions of the CF, provided in Chapter 5. Observe that, in terms of inference, ultimately, it is these Kolmogorov distances that we wish to make "small", by appropriate selection of the value of the truncation parameter c. Moreover, we wish to establish if the distribution of $\hat{X}$ is closer to the stable law than that of $X_{(0,c)}$ when allowing for the same computational cost, namely when both approximations are computed with the same parameter c.

The Chapter is organized as follows. We first present the preliminary study of the logarithm of the CF of the truncated PSR $X_{(0,c)}$. We then establish a bound on $\Delta(X, \hat{X})$, based on the smoothing lemma (5.35) and our bound on $\Delta(Z_{(c,\infty)}, Z)$ previously derived in Theorem 6.2.1. Subsequently, we adapt an existing literature result to obtain a bound on $\Delta(X, X_{(0,c)})$; observe that the latter deos not follow from the smoothing lemma, but directly considering the CDF divergence. However, we also compare the theoretical bounds in the smoothing lemma for both of these approximations, concluding that for the purposes of inference, the use of a truncated series together with an approximately Gaussian error term has superior statistical properties with respect to the simply truncated series, and it is likely a preferable choice in practice.

The results were first presented in Riabiz et al. [2018a, 2017c].

7.1 Study of the Function $q(u)$

When $W_j \sim \mathcal{N}(0, \sigma_W^2)$, the CF of the truncated PSR is given in (5.68). For convenience we write

$$\omega_{X_{(0,c)}}(u) = r(u)^c,$$

with

$$r(u) := \exp(-(1 - e^{-u} + u^a \Gamma(1 - a, u))). \tag{7.1}$$

Then, in this section, we study the function

$$q(u) := \log r(u) = -(1 - e^{-u} + u^a \Gamma(1 - a, u)). \tag{7.2}$$

Sign, Monotonicity and Asymptotics

It is clear that $q(0) = 0$ and $q(u) < 0$ for $u > 0$. In fact, using (B.18),

$$q'(u) = -au^{a-1}\Gamma(1 - a, u) \leq 0,$$

hence $q(u)$ is monotonically decreasing. Moreover , by (B.17), $q(u)$ is asymptotic to -1 at $u \to \infty$.

Slope as $u \to 0$

Since $a - 1 < 0$ and $\Gamma(1 - a, u)$ is decreasing in u we have that $q'(u)$ is monotonically increasing towards zero at $u \to \infty$, since that $\Gamma(1 - a, u)$ tends to $\Gamma(1 - a)$ as $u \to 0$. Note also that $q'(0)$ diverges to $-\infty$, which prevents a Taylor expansion around zero.

Concavity

We also compute

$$q''(u) = a((1 - a)u^{a-2}\Gamma(1 - a, u) + u^{-1}e^{-u}) \geq 0.$$

This is also monotonically decreasing to zero at $u \to \infty$, and in this case diverges to $+\infty$ at $u = 0$.

Bounds

These properties imply that $\log(\omega_{X_{(0,c)}}(u)) = cq(u)$ is *convex* and that it may be upper bounded by line segments. Indeed, the following construction will be useful in the proof of Theorem 7.2.1:[1] We will employ a piece-wise linear interpolating bound with N segments for $u \in [0,1]$, and a constant bound for $u > 1$. When $N = 1$, the bound is simply

$$\begin{aligned}\log(\omega_{X_{(0,c)}}(u)) = cq(u) \leq L^1(u) &:= \begin{cases} -cu((1-\exp(-1)) + \Gamma(1-a,1)), & u \in [0,1] \\ -c((1-\exp(-1)) + \Gamma(1-a,1)), & u > 1 \end{cases} \\ &= \begin{cases} m_0 u, & u \in [0,1] \\ k_{1,\infty}, & u > 1, \end{cases}\end{aligned}$$

where

$$\begin{aligned} m_0 = k_{1,\infty} &:= -c((1-\exp(-1)) + \Gamma(1-a,1)) \\ &= \log(\omega_{X_{(0,c)}}(1)) \\ &= \log(cq(1)) < 0. \end{aligned}$$

Bounding with more than one line segment in the domain $[0,1]$ turns out to be important for tightening the bound on $\Delta(X, \hat{X})$ and for capturing its dependence of the rate on α, as observed from simulating the bounds with numerical integration, see Section 7.2. When $N \geq 1$ we select N points u_i in $[0,1]$

$$0 =: u_0 < u_1 < \ldots < u_N := 1,$$

and the respective values of $\log(\omega_{X_{(0,c)}}(u))$

$$f_0 := 0 > f_1 := \log(\omega_{X_{(0,c)}}(u_1)) > \ldots > f_N := \log(\omega_{X_{(0,c)}}(u_N)).$$

The equation of the i-th line segment, for $i = 0, \ldots, N-1$, is $y = (m_i u + q_i)\mathbb{1}_{A_i}(u)$, with $A_i := [u_i, u_{i+1}]$ and

$$m_i := \frac{f_{i+1} - f_i}{u_{i+1} - u_i}, \qquad q_i := -m_i u_i + f_i.$$

[1] This is to preserve the same partition of the integration domain induced by $h(u)$ (6.21), and lead by the fact that $q(u)$ has *infinite slope in* $u = 0$.

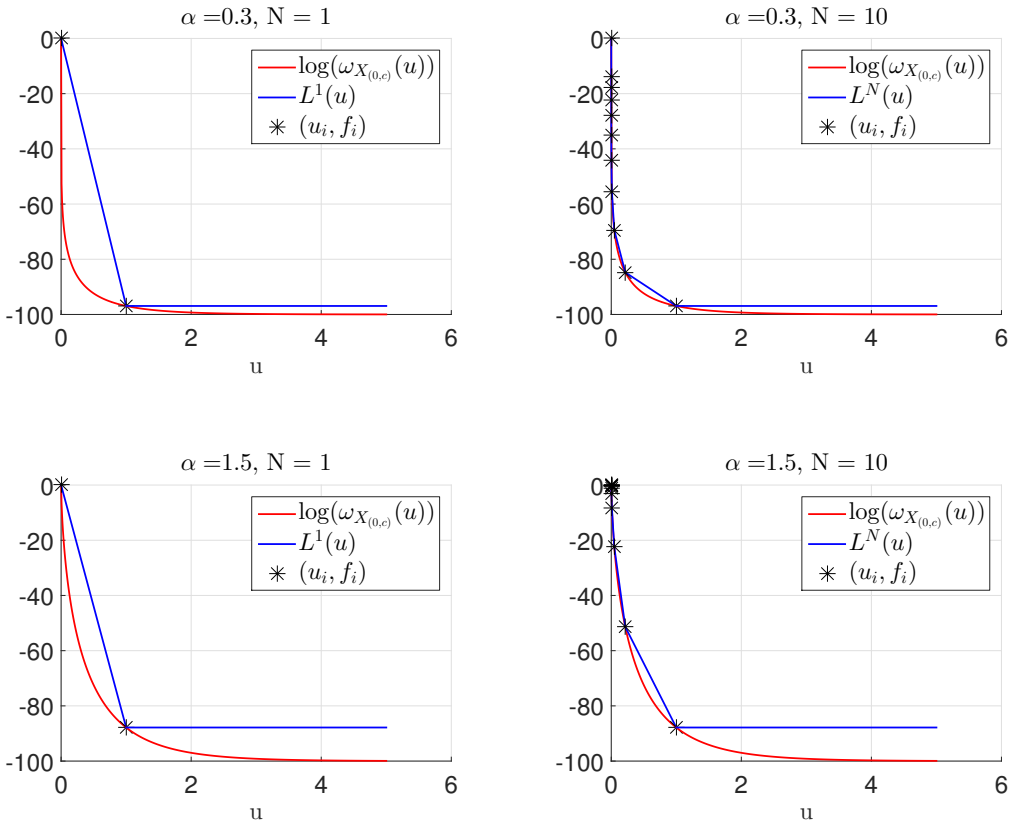


Figure 7.1 The red lines are examples of the function $\log(\omega_{X_{(0,c)}}(u))$, with $c = 100$ and $\alpha \in \{0.3, 1.5\}$. The blue lines represent the piece-wise linear bounds $L^1(u)$ and $L^N(u)$, with N = 10, and abscissae u_i logarithmically spaced on $[0, 1]$.

The general upper bound on $\log(\omega_{X_{(0,c)}}(u))$ then becomes

$$\log(\omega_{X_{(0,c)}}(u)) \leq L^N(u) := \begin{cases} \sum_{i=0}^{N-1}(m_i u + q_i)\mathbb{1}_{A_i}(u), & u \in [0,1] \\ k_{1,\infty}, & u > 1, \end{cases} \tag{7.3}$$

with $k_{1,\infty}$ as above.

Figure 7.1 shows examples of the function $\log(\omega_{X_{(0,c)}}(u)) = cq(u)$ and of its bounds $L^1(u)$ and $L^N(u)$, with N = 10, and abscissae u_i logarithmically spaced on $[0, 1]$, for two different values of α.

7.2 Nonasymptotic Bound on $\Delta(X, \hat{X})$

Here we establish a bound on $\Delta(X, \hat{X})$, based on the smoothing lemma (5.35) and the bound on the the distance of the PSR residual from a Gaussian, given in Theorem 6.2.1. Notice that in this case we deal with non-zero mean RVs (the mean is actually infinite for $\alpha < 1$). However, this does not constitute an issue in applying

the smoothing lemma, provided that we are able to provide finite bounds on the integral (5.35). Our result, presented in Riabiz et al. [2018a,b], is stated in the following.

Theorem 7.2.1. Let $\Delta(X, \hat{X})$ be the Kolmogorov distance between X and $\hat{X}$, as in (5.37), under the same assumptions and in the same notation as Theorem 6.2.1. Let $N \geq 1$, and introduce N arbitrary points

$$0 =: u_0 < u_1 < \cdots < u_N := 1,$$

together with the corresponding values of the logarithm of the CF $\omega_{X_{(0,c)}}(u)$ defined in (5.68)

$$f_0 := 0, \qquad f_i := \log(\omega_{X_{(0,c)}}(u_i)), \quad i = 1, 2, \ldots, N.$$

Also let,

$$m_i := \frac{f_{i+1} - f_i}{u_{i+1} - u_i}, \qquad q_i := -m_i u_i + f_i. \tag{7.4}$$

Then, for any $c > 1$, $\Delta(X, \hat{X})$ is bounded above by:

$$B_5(c, \alpha, N) := cK(a) \times \\ \times \left\{ \sum_{i=0}^{N-1} \frac{e^{q_i}}{\tilde{m}_i} \left[e^{\tilde{m}_i u_{i+1}} \left(u_{i+1} - \frac{1}{\tilde{m}_i} \right) - e^{\tilde{m}_i u_i} \left(u_i - \frac{1}{\tilde{m}_i} \right) \right] + \frac{e^{\tilde{k}_{(1,\infty)}}}{a \left(\tilde{l}_{(1,\infty)} \right)^{2/a}} \Gamma\left(\frac{2}{a}, \tilde{l}_{(1,\infty)} \right) \right\},$$

where

$$\begin{aligned} \tilde{m}_i &:= m_i + (c-1)\bar{g}, \\ k_{(1,\infty)} &:= -c((1 - \exp(-1)) + \Gamma(1-a, 1)) < 0, \\ \tilde{k}_{(1,\infty)} &:= k_{(1,\infty)} - (c-1)(e^{-1} - 1), \\ \tilde{l}_{(1,\infty)} &:= (c-1)\bar{\gamma}(a). \end{aligned} \tag{7.5}$$

Proof: Starting from the smoothing lemma (5.35), stating that $\Delta(X, \hat{X}) \leq \bar{I}(X, \hat{X})$, we proceed to bound $\bar{I}(X, \hat{X})$, defined as in (5.35). Using the expressions

(5.40) and (5.41), and performing the change of variables (5.57), we can write

$$\begin{aligned}\bar{I}(X,\hat{X}) &= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\left|\phi_{X_{(0,c)}}(s)\right|\left|\phi_{R_{(c,\infty)}}(s)-\phi_{\hat{R}}(s)\right|}{|s|}\,\mathrm{d}s \\ &= \frac{1}{\pi}\int_{0}^{\infty} \frac{\left|\omega_{X_{(0,c)}}(u)\right|\left|\omega_{R_{(c,\infty)}}(u)-\omega_{\hat{R}}(u)\right|}{u}\,\mathrm{d}u \\ &= \frac{1}{\pi}\int_{0}^{\infty} \frac{\left|\omega_{X_{(0,c)}}(u)\right|\left|\psi_{Z_{(c,\infty)}}(u)-\psi_{Z}(u)\right|}{u}\,\mathrm{d}u, \end{aligned} \tag{7.6}$$

where $\omega_{X_{(0,c)}}(u)$, $\omega_{R_{(c,\infty)}}(u)$ and $\omega_{\hat{R}}(u)$ are defined in (5.68), (5.67) and (5.70), respectively; the last equality follows again by (5.57), with $\psi_{Z_{(c,\infty)}}(w)$ as in (5.58) and $\psi_Z(w)$ in (5.63). The proof is in the following steps.

The term $\left|\psi_{Z_{(c,\infty)}}(u)-\psi_Z(u)\right|$ in the numerator of the integrand of (7.6) has already been bounded over the intervals $[0,1]$ and $(1,\infty)$ in the proof of Theorem 6.2.1, see Section 6.2. Combining equations (6.22) and (6.23), and recalling the definition of $K(a)$ (6.17), we obtain

$$\begin{aligned}\left|\psi_{Z_{(c,\infty)}}(u)-\psi_Z(u)\right| &\le cu^2\left(\frac{a}{2(2-a)}+\frac{1}{\eta^2}\right)\exp(-(c-1)h(u)) \\ &= cu^2\pi K(a)\exp(-(c-1)h(u)), \end{aligned} \tag{7.7}$$

with

$$h(u) = \begin{cases} -\bar{g}u, & u\in[0,1] \\ e^{-1}-1+u^a\bar{\gamma}(a), & u>1. \end{cases}$$

For the term $|\omega_{X_{(0,c)}}(u)|$ in (7.6), we recall the bounds obtained in Section 7.1 based on the fact that it is log-convex. From (7.3) and (7.7), the numerator of the integrand in (7.6) is bounded as,

$$\left|\omega_{X_{(0,c)}}(u)\right|\left|\psi_{Z_{(c,\infty)}}(u)-\psi_Z(u)\right| \le cu^2\pi K(a)\exp(\tilde{h}(u)),$$

with

$$\begin{aligned}\tilde{h}(u) &= \begin{cases} \sum_{i=0}^{N-1}\left((m_i+(c-1)\bar{g})u+q_i\right)\mathbb{1}_{A_i}(u), & u\in[0,1] \\ k_{1,\infty}-(c-1)(e^{-1}-1+u^a\bar{\gamma}(a)), & u>1 \end{cases} \\ &= \begin{cases} \sum_{i=0}^{N-1}\left(\tilde{m}_iu+q_i\right)\mathbb{1}_{A_i}(u), & u\in[0,1] \\ \tilde{k}_{1,\infty}-\tilde{l}_{1,\infty}u^a, & u>1, \end{cases}\end{aligned}$$

with m_i and q_i as in (7.4); $k_{1,\infty}$ as in (7.5), and where

$$\begin{aligned}\tilde{m}_i &:= m_i + (c-1)\bar{g},\\ \tilde{k}_{1,\infty} &= k_{1,\infty} - (c-1)(e^{-1}-1),\\ \tilde{l}_{1,\infty} &= (c-1)\bar{\gamma}(a).\end{aligned}$$

Finally, substituting into the integral (7.6)

$$\begin{aligned}\bar{I}(X,\hat{X}) &= cK(a)\int_0^\infty u\exp(-\tilde{h}(u))\,\mathrm{d}u\\ &= cK(a)(I_N^X(c)+J^X(c)),\end{aligned}$$

where $I_N^X(c)$ and $J^X(c)$ are the integrals on $(0,1)$ and on $(1,\infty)$, respectively. Computing such integrals we have

$$\begin{aligned}I_N^X(c) &= \sum_{i=0}^{N-1}\int_{A_i} u\exp(\tilde{m}_i u+q_i)\,\mathrm{d}u\\ &= \sum_{i=0}^{N-1}\exp(q_i)\int_{A_i} u\exp(\tilde{m}_i u)\,\mathrm{d}u\\ &= \sum_{i=0}^{N-1}\frac{\exp(q_i)}{\tilde{m}_i}\left[\exp(\tilde{m}_i u)\left(u-\frac{1}{\tilde{m}_i}\right)\right]_{u_i}^{u_{i+1}}\\ &= \sum_{i=0}^{N-1}\frac{e^{q_i}}{\tilde{m}_i}\left[e^{\tilde{m}_i u_{i+1}}\left(u_{i+1}-\frac{1}{\tilde{m}_i}\right)-e^{\tilde{m}_i u_i}\left(u_i-\frac{1}{\tilde{m}_i}\right)\right].\end{aligned}$$

Observe that, when $N=1$, $i=0$, $q_0=0$, $u_0=0$, $u_1=1$, $\tilde{m}_0 = m_0+(c-1)\bar{g} = \log(\omega_{X_{(0,c)}}(1))+(c-1)\bar{g}$, and the last equation becomes

$$\begin{aligned}I_1^X(c) &= \frac{1}{\tilde{m}_0}\left[e^{\tilde{m}_0}\left(1-\frac{1}{\tilde{m}_0}\right)+\frac{1}{\tilde{m}_0}\right]\\ &= \frac{1}{\tilde{m}_0}\left[e^{\tilde{m}_0}+\frac{1}{\tilde{m}_0}\left(1-e^{\tilde{m}_0}\right)\right].\end{aligned}$$

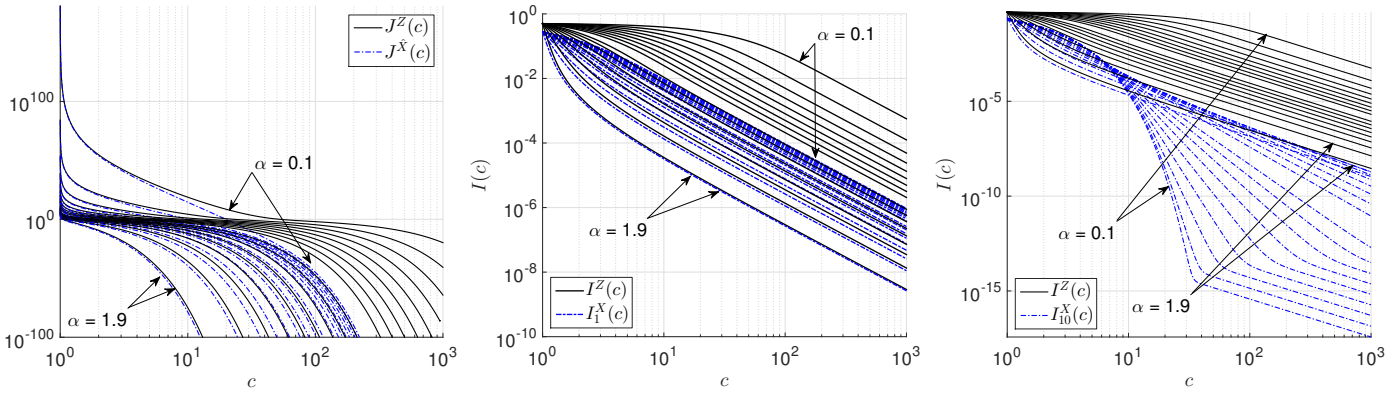


Figure 7.2 Comparison of the terms $I^Z(c)$ and $I_N^X(c)$, and $J^Z(c)$ and $J^X(c)$, for N = 1 and N = 10, and $\alpha = 0.1, 0.2, \ldots, 1.9$, plotted against $1 \leq c \leq 10^3$.

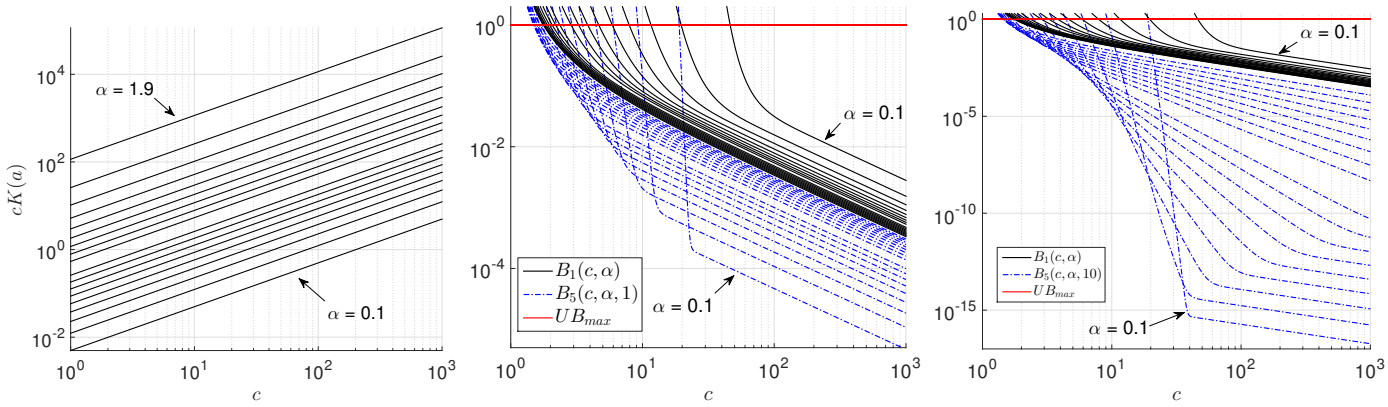


Figure 7.3 Left: functions $cK(a)$, with $a = \alpha/2$, plotted against $1 \leq c \leq 10^3$. Center and right: comparison of $B_4(c, \alpha) = cK(a)(I^Z(c) + J^Z(c))$ and $B_5(c, \alpha, N) = cK(a)(I_N^X(c) + J^X(c))$, for N = 1 and N = 10, and $\alpha = 0.1, 0.2, \ldots, 1.9$, plotted against $1 \leq c \leq 10^3$. The red horizontal lines are simply equal to 1, the maximum value of the Kolmogorov distance.

Finally,

$$
\begin{aligned}
J^X(c) &= \int_1^\infty u \exp\left(\tilde{k}_{1,\infty} - \tilde{l}_{1,\infty} u^a\right) \mathrm{d}u \\
&= \exp\left(\tilde{k}_{1,\infty}\right) \int_1^\infty u \exp\left(-\tilde{l}_{1,\infty} u^a\right) \mathrm{d}u \\
&= \frac{\exp\left(\tilde{k}_{1,\infty}\right)}{a\left(\tilde{l}_{1,\infty}\right)^{2/a}} \int_{\tilde{l}_{1,\infty}}^\infty t^{\frac{2}{a}-1} \exp(-t)\, \mathrm{d}t \\
&= \frac{\exp\left(\tilde{k}_{1,\infty}\right)}{a\left(\tilde{l}_{1,\infty}\right)^{2/a}} \Gamma\left(2/a, \tilde{l}_{1,\infty}\right).
\end{aligned}
$$

Combining the above results leads to the statement.

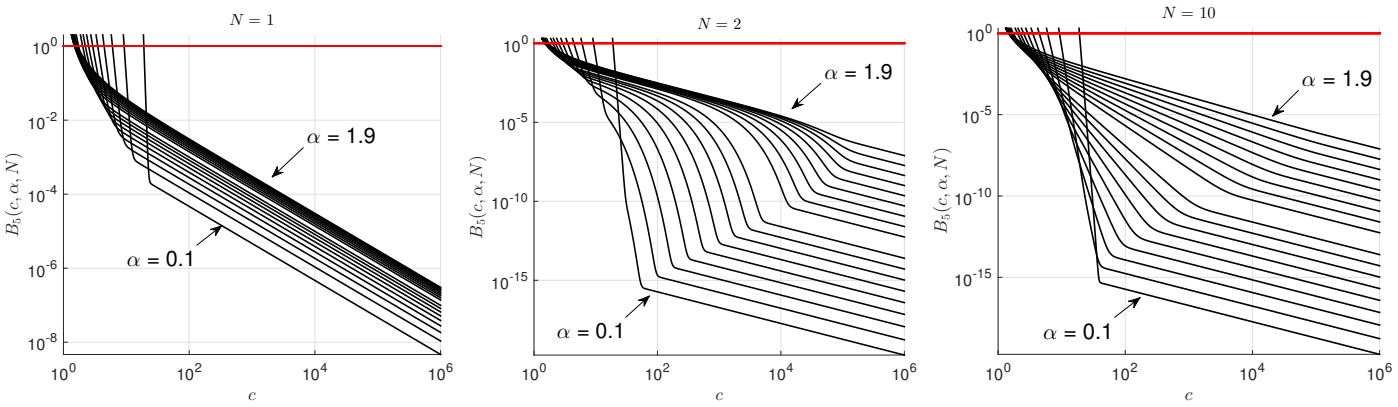


Figure 7.4 Bounds on $\Delta(X, \hat{X})$. Each curve shows the bound $B_5(c, \alpha, N)$ for $\alpha = 0.1, 0.9, \ldots, 1.9$, $\alpha \neq 1$, plotted against $1 \leq c \leq 10^6$. The N abscissae $u_0, \ldots, u_{N-1}$ are logarithmically spaced on the interval [0,1]. The red horizontal lines are simply equal to 1, the maximum value of the Kolmogorov distance.

Remark 10. Observe that the bounds $I_N^X(c)$ and $J^X(c)$ are smaller than $I^Z(c)$ and $J^Z(c)$,[2] because the latter correspond to the former when

$$\omega_{X_{(0,c)}}(u) \equiv 1.$$

Figure 7.2 shows a comparison between these quantities plotted against c, for a number of values of α, and $N = 1$ or $N = 10$ (notice that the choice of N affects only $I_N^X(c)$, not $J^X(c)$). We observe that

(i) both $J^Z(c)$ and $J^X(c)$ have asymptotic rates dependent on α and that larger values of α give smaller bounds;

(ii) on the other hand, $I^Z(c)$ and $I_N^{\hat{X}}(c)$ have asymptotic rate independent on α, but their nonasymptotic behaviour does depend on α for a wide range of values of c;

(iii) moreover, for $N > 1$, lower values of α now give lower values of the bound $I_N^X(c)$, whereas for $N = 1$, the dependence of $I_N^X(c)$ on α is the same as that of $I^Z(c)$. But even in this case, the dependence of $B_5(c, \alpha, N)$ of α is the opposite to that of $B_1(c, \alpha)$, because of the coefficient $cK(a)$; this is illustrated in Figure 7.3.

■

The N abscissae u_i and ordinates f_i serve to define a piece-wise linear envelope on $\omega_{X_{(0,c)}}(u)$ for $u \in [0, 1]$, that is used in the proof, as explained also in Section 7.1.

[2]For the same reason, we can state that $\bar{I}(X, \hat{X}) \leq \bar{I}(Z_{(c,\infty)}, Z)$, and that $B_5(c, \alpha, N) \leq B_1(c, N)$.

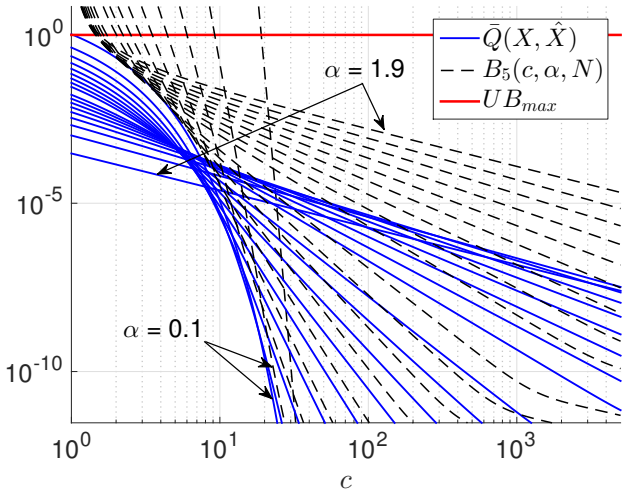


Figure 7.5 Bounds on $\Delta(X, \hat{X})$. The blue solid lines represent $\bar{Q}(X, \hat{X})$; the black dashed lines represent $B_5(c, \alpha, N)$ with $N = 10$ and logarithmically spaced points. The values are plotted for $\alpha = 0.1, 0.2, \ldots, 1.9$, $\alpha \neq 1$, and for $1 \leq c \leq 5000$. The red horizontal line is simply equal to 1, the maximum value of the Kolmogorov distance.

Increasing N improves (i.e., decreases) $B_5(c, \alpha, N)$, but the changes are minimal for $N \geq 10$ and logarithmically spaced points, as shown in Figure 7.4, where three different values of $N = 1$, 2, 10 are compared.

In Figure 7.5 we compare the numerical estimates $\bar{Q}(X, \hat{X})$ of $\bar{I}(X, \hat{X})$, first presented in Riabiz et al. [2017c], with the bound $B_5(c, \alpha, N)$ for $N = 10$. Note that the bound of Theorem 7.2.1 correctly captures the dependence on α, and that the approximation error is lower for smaller values of α, a reversal of the trend shown in Figure 6.3. We believe this is because, as α decreases, the relative significance of the residual term is much smaller, when compared with the heavy-tailed initial terms in the PSR. We also observe that the rate of convergence is dramatically better for smaller α, again in contrast with the analysis of the residual approximation in Figure 6.3. Finally, it seems that $B_5(c, \alpha, N)$ has the same asymptotic behaviour as $B_1(c, \alpha)$ for $c \to \infty$, but these two bounds have reversed asymptotic ordering with respect to α; see also Remark 10 in the proof of Theorem 7.2.1. The numerical simulations become unstable when $c \to \infty$, and this is why in the figure we are choosing the range $1 \leq c \leq 5000$, in which the error bands of the numerical integrals (not displayed here) are negligibly small.

7.3 Nonasymptotic Bound on $\Delta(X, X_{(0,c)})$

The following bound on $\Delta(X, X_{(0,c)})$ (5.38), is an adaptation of the result of Ledoux and Paulauskas [1996]. Its proof is based on direct computations and does not rely on the smoothing lemma.

Proposition 1. Let $\Delta(X, X_{(0,c)})$ be defined as in (5.38). Under the same assumptions and in the same notation as in Theorem 6.2.1

$$\Delta(X, X_{(0,c)}) \leq B_6(c, \alpha) := \frac{\exp(-1/2)}{\sqrt{(2\pi)}} \times \sqrt{\Gamma\left(\frac{\alpha+4}{\alpha}\right)\left(\frac{\alpha}{4-\alpha}c^{\frac{\alpha-4}{\alpha}} + \left(\frac{\alpha}{2-\alpha}c^{\frac{\alpha-2}{\alpha}}\right)^2\right)}. \tag{7.8}$$

Proof: We adapt the result of Ledoux and Paulauskas [1996] on the convergence of the truncated PSR with $W_1 \sim \mathcal{N}(0, \sigma_W^2 = 1)$. The difference here is that the number of terms in $X_{(0,c)}$ is random, and that we allow the variance σ_W^2 to not necessarily be equal to 1. Then for $x > 0$

$$\begin{aligned}
\Delta(X, X_{(0,c)}) &= \sup_{x\in\mathbb{R}} \left| F_X(x) - F_{X_{(0,c)}}(x) \right| \\
&= \sup_{x\in\mathbb{R}} \left| \mathbb{P}(X \leq x) - \mathbb{P}(X_{(0,c)} \leq x) \right| \\
&= \sup_{x\in\mathbb{R}} \left| \mathbb{E}[\mathbb{1}(X \leq x)] - \mathbb{E}[\mathbb{1}(X_{(0,c)} \leq x)] \right|.
\end{aligned}$$

Now, using the law of total expectation, and conditioning first on the number $N_{(0,c)}$ of terms in the truncated PSR[3] and the Poisson arrival times $\{\Gamma_j\}$, we can write

$$\begin{aligned}
\mathbb{E}[\mathbb{1}\left(X \leq x\right)] &= \mathbb{E}\left[\mathbb{1}\left(\sum_{j=1}^{\infty} \Gamma_j^{-1/\alpha} W_j \leq x\right)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}\left(\sum_{j=1}^{\infty} \Gamma_j^{-1/\alpha} W_j \leq x\right) \middle| N_{(0,c)}, \{\Gamma_j\}_{j=1}^{\infty}\right]\right] \\
&= \mathbb{E}\left[\mathbb{P}\left(\sum_{j=1}^{\infty} \Gamma_j^{-1/\alpha} W_j \leq x \middle| N_{(0,c)}, \{\Gamma_j\}_{j=1}^{\infty}\right)\right] \\
&= \mathbb{E}\left[F_Z\left(\frac{x}{\sigma_W S}\right)\right],
\end{aligned}$$

[3] Recalling the definition of the truncated PSR (5.2), with obvious notation $N_{(0,c)} \sim \text{Poisson}(c)$ denotes the number of terms in $X_{(0,c)}$.

where $F_Z(\cdot)$ is the CDF of the standard normal distribution, and S^2 is defined in (2.37). Notice that in the above expression we could avoid conditioning on $N_{(0,c)}$, because X is the full PSR, and it is thus independent on how many Γ_j are lower than c. However, accounting for $N_{(0,c)}$ in the notation is necessary for computing the CDF of $X_{(0,c)}$, that is similarly computed

$$\mathbb{E}[\mathbb{1}_{(X_{(0,c)} \leq x)}] = \mathbb{E}\left[F_Z\left(\frac{x}{\sigma_W S_{N_{(0,c)}}}\right)\right],$$

where

$$S^2_{N_{(0,c)}} := \sum_{j=1}^{N_{(0,c)}} \Gamma_j^{-2/\alpha}, \tag{7.9}$$

with the convention that $S^2_{N_{(0,c)}} = 0$ if $N_{(0,c)} = 0$, and $F_Z(-\infty) = 0$, $F_Z(\infty) = 1$. Then, given that $\Gamma_j > 0$ for any $j \geq 1$ with probability 1, it follows that $S^2_{N_{(0,c)}} < S^2$, and, for $x > 0$

$$\Delta(X, X_{(0,c)}) = \sup_{x \in \mathbb{R}} \mathbb{E}\left[F_Z\left(\frac{x}{\sigma_W S_{N_{(0,c)}}}\right) - F_Z\left(\frac{x}{\sigma_W S}\right)\right].$$

The argument of the expectation above can be bounded as

$$\begin{aligned}
F_Z\left(\frac{x}{\sigma_W S_{N_{(0,c)}}}\right) - F_Z\left(\frac{x}{\sigma_W S}\right) &= \frac{1}{\sqrt{(2\pi)}} \int_{\frac{x}{\sigma_W S}}^{\frac{x}{\sigma_W S_{N_{(0,c)}}}} \exp\left(-u^2/2\right) \mathrm{d}u \\
&\leq \frac{x}{\sigma_W \sqrt{(2\pi)}} \left(\frac{1}{S_{N_{(0,c)}}} - \frac{1}{S}\right) \exp\left(-\frac{x^2}{2\sigma_W^2 S^2}\right) \\
&\leq \frac{1}{\sigma_W \sqrt{(2\pi)}} \left(\frac{S - S_{N_{(0,c)}}}{S_{N_{(0,c)}} S}\right) S\sigma_W \exp(-1/2) \\
&= \frac{\exp(-1/2)}{\sqrt{(2\pi)}} \left(\frac{S^2 - S_{N_{(0,c)}} S}{S_{N_{(0,c)}} S}\right) \\
&\leq \frac{\exp(-1/2)}{\sqrt{(2\pi)}} \left(\frac{S^2 - S^2_{N_{(0,c)}}}{S^2_{N_{(0,c)}}}\right),
\end{aligned}$$

where in the first inequality we have used the constant bound of the integrand on the integration domain, while in the second inequality we have used the fact that the mode of the Rayleigh distribution is achieved for $x = \sigma_W S$; finally, in the third

inequality we have used (twice) the fact that $S^2_{N_{(0,c)}} < S^2$. Notice that this upper bound does not depend on σ_W. Finally, taking the expectation,

$$\begin{aligned}
\Delta(X, X_{(0,c)}) &\leq \frac{\exp(-1/2)}{\sqrt{(2\pi)}} \mathbb{E}\left[S^{-2}_{N_{(0,c)}}(S^2 - S^2_{N_{(0,c)}})\right] \\
&\leq \frac{\exp(-1/2)}{\sqrt{(2\pi)}} \left(\mathbb{E}\left[S^{-4}_{N_{(0,c)}}\right]\right)^{1/2} \left(\mathbb{E}\left[(S^2 - S^2_{N_{(0,c)}})^2\right]\right)^{1/2} \\
&\leq \frac{\exp(-1/2)}{\sqrt{(2\pi)}} \left(\mathbb{E}\left[\Gamma_1^{4/\alpha}\right]\right)^{1/2} \left(\mathbb{E}\left[(S^2 - S^2_{N_{(0,c)}})^2\right]\right)^{1/2} \\
&= \frac{\exp(-1/2)}{\sqrt{(2\pi)}} \left(\Gamma\left(\frac{\alpha+4}{\alpha}\right)\right)^{1/2} \times \left(\frac{\alpha}{4-\alpha} c^{\frac{\alpha-4}{\alpha}} + \left(\frac{\alpha}{2-\alpha} c^{\frac{\alpha-2}{\alpha}}\right)^2\right)^{1/2},
\end{aligned}$$

where in the second inequality we have used the Cauchy-Schwartz inequality; in the third inequality we have used that $S^2_{N_{(0,c)}} > \Gamma_1^{-2/\alpha}$, where Γ_1 is the smallest of the Γ_j variables, hence exponentially distributed. Thus

$$\mathbb{E}\left[\Gamma_1^{4/\alpha}\right] = \int_0^\infty u^{(4/\alpha+1)-1} e^{-u}\, \mathrm{d}u = \Gamma\left(\frac{4}{\alpha}+1\right), \tag{7.10}$$

where $\Gamma(\cdot)$ is the gamma function (B.12). On the other hand, $S^2 - S^2_{N_{(0,c)}} = \sum_{j=N_{(0,c)}+1}^\infty \Gamma_j^{-2/\alpha} = \lim_{d\to\infty} \sum_{j:\Gamma_j\in(c,d)} \Gamma_j^{-2/\alpha}$, and we know that such limit exists through the PSR, being the second moment of the PSR residual with deterministic $W_j = 1$. Hence

$$\begin{aligned}
\mathbb{E}\left[(S^2 - S^2_{N_{(0,c)}})^2\right] &= \lim_{d\to\infty} \mathbb{E}\left[\sum_{j:\Gamma_j\in(c,d)} \Gamma_j^{-2/\alpha} \sum_{i:\Gamma_i\in(c,d)} \Gamma_i^{-2/\alpha}\right] \\
&= \lim_{d\to\infty} \mathbb{E}\left[\mathbb{E}\left[\sum_{j=1}^{N_{(c,d)}} \Gamma_j^{-2/\alpha} \sum_{i=1}^{N_{(c,d)}} \Gamma_i^{-2/\alpha} \middle| N_{(c,d)}\right]\right] \\
&= \lim_{d\to\infty} \mathbb{E}\left[\mathbb{E}\left[\sum_{j=1}^{N_{(c,d)}} U_j^{-2/\alpha} \sum_{i=1}^{N_{(c,d)}} U_i^{-2/\alpha} \middle| N_{(c,d)}\right]\right] \\
&= \lim_{d\to\infty} \mathbb{E}[N_{(c,d)}]\mathbb{E}[U_1^{-4/\alpha}] + \mathbb{E}[N^2_{(c,d)} - N_{(c,d)}](\mathbb{E}[U_1^{-2/\alpha}])^2 \\
&= \frac{\alpha}{4-\alpha} c^{\frac{\alpha-4}{\alpha}} + \left(\frac{\alpha}{2-\alpha} c^{\frac{\alpha-2}{\alpha}}\right)^2,
\end{aligned} \tag{7.11}$$

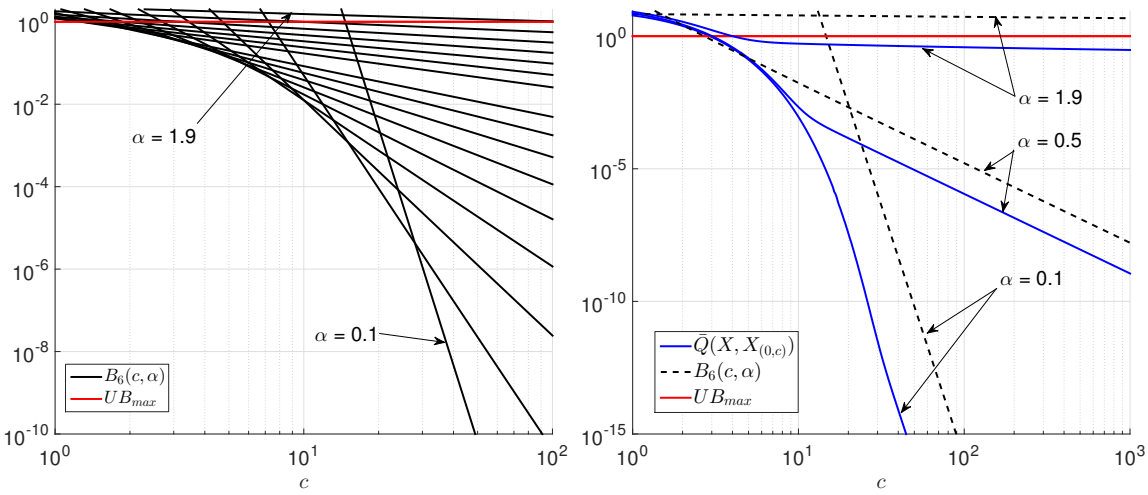


Figure 7.6 Bounds on $\Delta(X, X_{(0,c)})$. Left: each curve represents the values of the bound $B_6(c, \alpha)$ for $\alpha = 0.1, 0.9, \ldots, 1.9$, $\alpha \neq 1$, plotted against $1 \leq c \leq 10^2$. Right: the blue solid lines represent $\bar{Q}(X, X_{(0,c)})$; the black dashed lines represent $B_6(c, \alpha)$. The values are plotted for α $in\{0.1, 0.5, 1.9\}$, and for $1 \leq c \leq 1000$. The red horizontal lines are simply equal to 1, the maximum possible value of the Kolmogorov distance.

where $\{U_j\}$ are RVs uniformly distributed on (c, d), as in (5.7).

■

From (7.8), it is easy to see that $B_6(c, \alpha) = O(c^{(\alpha-2)/\alpha})$, and a corresponding lower bound of the same asymptotic order is established in Ledoux and Paulauskas [1996]. The left-hand side of Figure 7.6 shows the bound $B_6(c, \alpha)$ for a range of values of α, while the right-hand side compares it to the numerical estimates $\bar{Q}(X, X_{(0,c)})$ of $\bar{I}(X, X_{(0,c)})$ as in (5.35). Notice that, in general, we will have that

$$\Delta(X, X_{(0,c)}) \leq \min\{1,\ B_6(c, \alpha),\ \bar{I}(X, X_{(0,c)})\},$$

and, in fact, In the figure there is no (uniform with respect to c and α) dominance between $\bar{Q}(X, X_{(0,c)})$ and $B_6(c, \alpha)$. However $B_6(c, \alpha)$ and $\bar{Q}(X, X_{(0,c)})$ appear to have the same asymptotic order.

Observe that none of the bounds on $\Delta(X, \hat{X})$ and $\Delta(X, \mathbf{X}_{(0,c)})$ that we established depends on σ_W. This is consistent with the considerations in Remark 8.

7.4 Comparison of $\bar{I}(X, \hat{X})$ and $\bar{I}(X, X_{(0,c)})$

Finally, we provide a result that compares the bounds on $\Delta(X, \hat{X})$ and $\Delta(X, X_{(0,c)})$ that derive from the smoothing lemma. Our result indicates that, in the symmetric case $W_1 \sim \mathcal{N}(\mu_W, \sigma_W^2)$, adding a Gaussian approximation of the residual to the

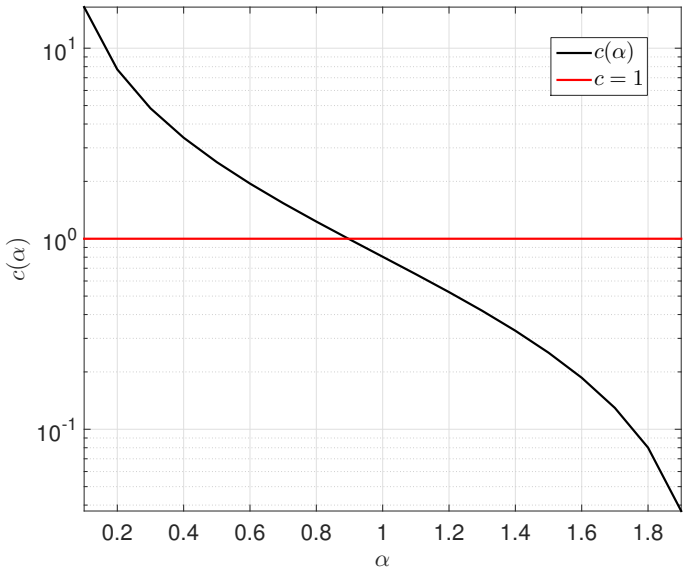


Figure 7.7 The function $c(\alpha)$ such that $\bar{I}(X, \hat{X}) < \bar{I}(X, X_{(0,c)})$ for $c > c(\alpha)$ in the symmetric stable distribution, plotted against $\alpha = 0.1, \ldots, 1.9$. The red horizontal line corresponds to $c = 1$, that is the minimum value of c that we consider relevant for practical purposes.

truncated PSR is likely to provide a better approximation to the α-stable distribution than the PSR alone, for most values of the truncation parameter c. In fact, the following lemma holds.

Proposition 2. Let $\bar{I}(X, \hat{X})$ and $\bar{I}(X, X_{(0,c)})$ be defined as in (5.35), under the same assumptions and notation as in Theorem 6.2.1. Then, for any $\alpha \in (0, 2)$, $\alpha \neq 1$, we have that

$$\bar{I}(X, \hat{X}) < \bar{I}(X, X_{(0,c)})$$

for all

$$c > c(\alpha) := \frac{\log(2)}{\gamma(1 - \alpha/2, 1) + e^{-1} - 1}. \tag{7.12}$$

Proof: Recall that $\bar{I}(X, \hat{X})$ is given by (7.6). For $\bar{I}(X, X_{(0,c)})$, using (5.40) and (5.41), and performing the change of variables (5.57), we similarly have,

$$\bar{I}(X, X_{(0,c)}) = \frac{1}{\pi} \int_0^\infty \frac{\left|\omega_{X_{(0,c)}}(u)\right| \left|\psi_{Z_{(c,\infty)}}(u) - 1\right|}{u} \, \mathrm{d}u. \tag{7.13}$$

We proceed by comparing the integrands in (7.6). and (7.13). We aim at finding a function $c(\alpha)$ s.t., for all $c > c(\alpha)$ and $u > 0$,

$$\left|\psi_{Z_{(c,\infty)}}(u) - \psi_Z(u)\right| < \left|\psi_{Z_{(c,\infty)}}(u) - 1\right|, \tag{7.14}$$

where $\log(\psi_Z(u)) = -cu/\eta$, as in (5.63), and $\psi_{Z_{(c,\infty)}}$, defined in (5.58), satisfies,

$$\begin{aligned}\log(\psi_{Z_{(c,\infty)}}(u)) &= c(1-\exp(-u)-u^a\gamma(1-a,u)) \\ &= cg(u),\end{aligned}$$

with $g(\cdot)$ defined in (6.3). Using the fact that $g(u) < 0$, see (6.4), we have that $|\omega_{R_{(c,\infty)}}(u) - 1| = 1 - \omega_{R_{(c,\infty)}}(u)$. Furthermore, using (6.10), we have $|\omega_{R_{(c,\infty)}}(u) - \omega_{\hat{R}}(u)| = \omega_{R_{(c,\infty)}}(u) - \omega_{\hat{R}}(u)$, and (7.14) becomes

$$v(u) := 1 - 2\exp(cg(u)) + \exp(-cu/\eta) > 0.$$

Based on (6.5), we have that, for $u \in (0,1]$

$$\begin{aligned}v(u) &> (1 - 2\exp(-cu/2\eta) + \exp(-cu/\eta)) \\ &= (1-\exp(-cu/2\eta))^2 > 0,\end{aligned}$$

hence (7.14) holds for any $c > 0$ when $u \leq 1$. We then consider the case $u > 1$. By monotonicity of $g(\cdot)$, see (6.9), we have that $g(u) < g(1)$ for $u > 1$, leading to

$$v(u) > 1 - 2\exp(cg(1)) + \exp(-uc/\eta) := z(u), \quad u > 1,$$

where $z(u)$ is a lower bound on $v(u)$ for $u > 1$. We know that $z(1) = v(1) > 0$ as shown above, and also

$$z'(u) = -c/\eta \exp(-uc/\eta) < 0,$$

implying that the lower bound $z(u)$ on $v(u)$ is decaying. Furthermore, $\lim_{u\to\infty} z(u) = 1 - 2\exp(cg(1))$, so that, for all $u \geq 1$,

$$v(u) \geq 1 - 2\exp(cg(1)),$$

and the right-hand side above is positive as long as $c > -\log(2)/g(1) = c(\alpha)$, as required (recall that $g(1) < 0$ and that $a = \alpha/2$).

■

Proposition 2 suggests that the Gaussian residual approximation produces a smaller approximation error than simply truncating the series, a result borne out by previous experimental simulations also for the asymmetric case, see [Lemke, 2014, p. 56-57], and Figure 5.2.

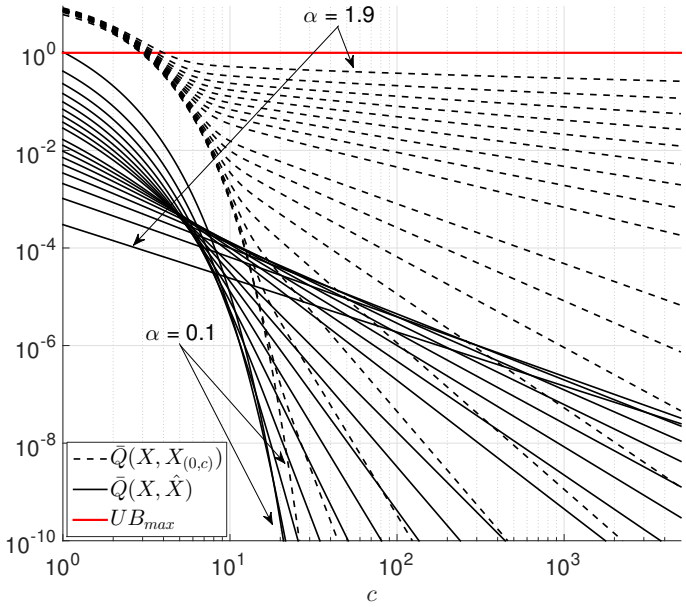


Figure 7.8 Numerical bounds on $\Delta(X, \hat{X})$ and on $\Delta(X, X_{(0,c)})$. The solid lines correspond to $\bar{Q}(X, \hat{X})$ and the dashed lines to $\bar{Q}(X, X_{(0,c)})$ for $\alpha = 0.1, 0.2, \ldots, 1.9$, $\alpha \neq 1$, plotted against $1 \leq c \leq 5000$. The red horizontal line is simply equal to 1, the maximum value of the Kolmogorov distance.

The values of $c(\alpha)$ are plotted in Figure 7.7. Recalling that $\gamma(1-a, 1) = \int_0^1 t^{-a} e^{-t}\, \mathrm{d}t$, we have

$$\lim_{a \to 0} \gamma(1-a, 1) := \int_0^1 e^{-t}\, \mathrm{d}t = 1 - e^{-1},$$
$$\lim_{a \to 1} \gamma(1-a, 1) := \int_0^1 t^{-1} e^{-t}\, \mathrm{d}t = +\infty,$$

where the second integral is the limit, as $x \to 0$, of the exponential integral $E_1(x)$, defined for example in Olver et al. [2018]. Hence $c(\alpha)$ diverges for $\alpha \to 0$, and converges to 0 for $\alpha \to 2$ (in particular, $c(\alpha) < 1$ for $\alpha > 1$, even if, for practical purposes, we can consider that $c > 1$ always, and this corresponds to the red line in the figure).

Remark 11. Observe that $c(\alpha) < 16$ for all $\alpha > 0.1$. Moreover, the condition $c > c(a)$ in the proposition is only sufficient, and, in fact, the numerical estimates of the integrals $\bar{I}(X, \hat{X})$ and $\bar{I}(X, X_{(0,c)})$ are s.t.

$$\bar{Q}(X, \hat{X}) < \bar{Q}(X, X_{(0,c)}), \qquad \forall c > 1,$$

as shown in Figure 7.8. This is an indication that the condition $c > c(\alpha)$ could be removed through more detailed considerations.

Table 7.1 Summary of the bounds given in Theorems 6.2.1 and 7.2.1, Propositions 1 and 2, and Remarks 10 and 11 (red asterisks), and of the numerical simulations shown in Figures 6.3, 7.5 and 7.8 (approximation symbols with blue squares).

Theoretical bounds	$B_1(c,\alpha)$	$\overset{*}{\geq}$	$B_5(c,\alpha,N)$		
	$\wr\vee$ *		$\wr\vee$ *		
Smoothing lemma bounds	$\bar{I}(Z_{(c,\infty)}, Z)$	$\overset{*}{\geq}$	$\bar{I}(X, \hat{X})$	$\overset{*}{\leq}$ [if $c<c(\alpha)$]	$\bar{I}(X, X_{(0,c)})$
	$\wr\wr$ ▪		$\wr\wr$ ▪		$\wr\wr$ ▪
Numerical bounds	$\bar{Q}(Z_{(c,\infty)}, Z)$		$\bar{Q}(X, \hat{X})$	$\overset{*}{<}$	$\bar{Q}(X, X_{(0,c)})$
	$\wr\vee$		$\wr\vee$		$\wr\vee$
Kolmogorov distances	$\Delta(Z_{(c,\infty)}, Z)$		$\Delta(X, \hat{X})$		$\Delta(X, X_{(0,c)})$
					$\wr\wedge$ *
Direct theoretical bounds					$B_6(c,\alpha)$

For easy reference, Table 7.1 provides a summary of the analytical and numerical bounds that were established in Chapters 5 and 7, when approximating the PSR of symmetric α-stable RVs. Observe that, for simplicity, we are reporting only the bounds related to the smoothing lemma with smoothing parameter set to infinity.

Chapter 8

Approximation of the PSR for Multivariate Stable Stochastic Integrals

In this Chapter we provide some results on a LV representation for stochastic integrals driven by an α-stable Lévy process. As anticipated in Section 2.3, the stochastic integral (SI) is either a stable RV or a stable random vector. Our focus here is again on Poisson series representations, because they allow us to deal with stable distributions with arbitrary skeweness or spectral measure. In particular, we consider the modified PSR (MPSR) that has recently been introduced by Lemke and Godsill [2011] and Lemke [2014], and that transforms the PSR for stochastic integrals in a way that yields a conditionally Gaussian representation of the corresponding stable distribution. As explained in Section 1.3, the advantage of such a framework is to enable for efficient Bayesian state inference (Rao-Blackwellized paricle filters). The approximate MPSR in the above cited literature is focussed on scalar stochastic integrals, and our contribution is to generalize such approximation to the multivariate case .

The structure of the Chapter is as follows. First, in Section 8.1 we recall the PSR and MPSR for stochastic integrals, with an emphasis on the vector notation. These are infinite series that need to be truncated, as for the case of random variables. In Section 8.2 we formulate both the GAA and the GAMA schemes for the multivariate MPSR residual. This result is then used in Section 8.3 to approximately simulate the solution of the SDE describing the linear continuous-time model (2.16), for which we also provide an experimental validation of our approximation. The latter is conceptually similar to the Monte Carlo simulations in Section 5.3.1, aimed at testing

the Gaussianity of the PSR residual, and it will be the subject of future analytical studies.

Part of the results from this chapter was presented in Riabiz and Godsill [2017].

8.1 Series Representations for Stable Stochastic Integrals

Up to this point of the thesis, we have dealt with the PSR of α-stable random variables, given in equation (2.29). In [Samoradnitsky and Taqqu, 1994, Sections 3.9, 3.10], a PSR is provided also for random measures $M(\cdot)$, and thus extended to the corresponding stochastic integrals, with expression as in (2.17)-(2.18)

$$\mathbf{I}(\mathbf{f}) = \int_E \mathbf{f}(u)\,\mathrm{d}\ell(u) = \int_E \mathbf{f}(u)M(\mathrm{d}u),$$

with $\ell(t) - \ell(s) \sim \mathcal{S}_\alpha((t-s)^{1/\alpha}, \beta, 0)$ as in (2.12) and $\mathbf{f}$ a d-dimensional function such that (2.19) holds. We remark that, for our purposes of time series modelling, we consider integrals wrt Lévy processes, so that $M(\cdot)$ is a stable random measure with control measure $m(\cdot)$ the Lebesgue measure, and constant skeweness intensity $\beta(u) = \beta$. However, the PSR reported in the following section is formulated for more generic stable random measures and stochastic integrals (driven by a general stable process).

8.1.1 PSR for Stable Stochastic Integrals

Let $\{\Gamma_j\}_{j=1}^\infty$ be a sequence of arrival times of a unite rate Poisson process, as before, and let $\{(V_j, \gamma_j)\}_{j=1}^\infty$ be a sequence of i.i.d. random vectors, independent of $\{\Gamma_j\}_{j=1}^\infty$, such that

$$V_j \sim \hat{m} := \frac{m}{m(E)},$$
$$\mathbb{P}(\gamma_j = 1|V_j) = 1 - \mathbb{P}(\gamma_j = -1|V_j) = \frac{1+\beta(V_j)}{2}.$$

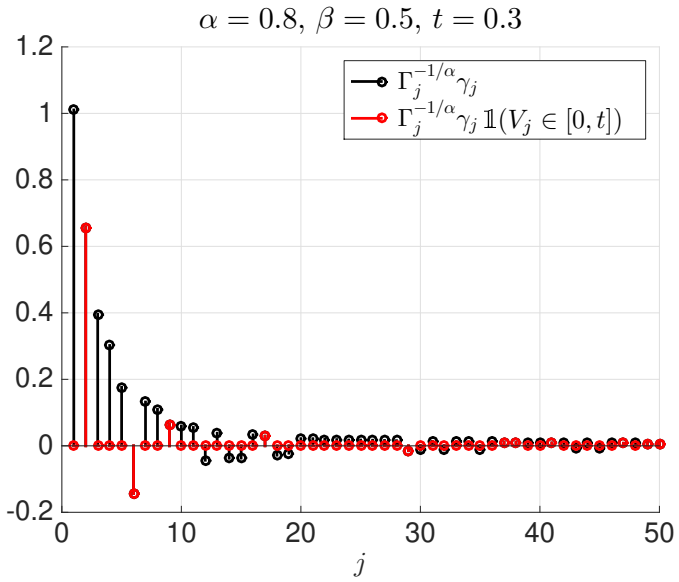


Figure 8.1 Red: Summands $\Gamma_j^{-1/\alpha}\gamma_j \mathbb{1}(V_j \in [0,t])$ of the PSR realization for the stable random measure of $M(0,t) = \ell(t)$, with $\alpha = 0.8$, $\beta(u) = \beta = 0.5$, $t = 0.3$, and time horizon $T = 1$. The black bars correspond to the summands of the same PSR realization for the stable random measure of $[0,T]$.

Then, for the stochastic integral $\mathbf{I}(\mathbf{f})$ in (2.18), the following equality in distribution holds: $\mathbf{I}(\mathbf{f}) \stackrel{\mathcal{D}}{=} \mathbf{S}_1(\mathbf{f})$, where $\mathbf{S}_1(\mathbf{f})$ is the a.s. convergent random series

$$\mathbf{S}_1(\mathbf{f}) := \begin{cases} (C_\alpha m(E))^{1/\alpha} \sum_{j=1}^{\infty} \left[\gamma_j \Gamma_j^{-1/\alpha} \mathbf{f}(V_j) - b_j^{(\alpha)} \int_E \mathbf{f}(u)\beta(u)\hat{m}(\mathrm{d}u)\right] & \text{if } \alpha \neq 1, \\ \frac{2}{\pi} m(E) \sum_{j=1}^{\infty} \left[\gamma_j \Gamma_j^{-1/\alpha} \mathbf{f}(V_j) - b_j^{(\alpha)} \int_E \mathbf{f}(u)\beta(u)\hat{m}(\mathrm{d}u)\right] + \eta_{\mathbf{f}} & \text{if } \alpha = 1, \end{cases} \tag{8.1}$$

with C_α as in (2.35), b_j as in (2.32) and $\eta_{\mathbf{f}}$ a function of $\mathbf{f}$ non-zero only if $\alpha = 1$, a case that we do not consider in the following (the expression can be found in [Samoradnitsky and Taqqu, 1994, p]).

For example, the red bars in Figure 8.1 are the first few summands of the PSR of the stable Lévy process at time $0 \leq t \leq T$, corresponding to the stochastic integral $I(\mathbb{1}([0,t])) = M([0,t]) = \ell(t)$, with $t = 0.3$, in a case where $\alpha < 1$, so that the deterministic part of (8.1) is zero (we are also omitting the multiplicative constant $(C_\alpha m(E))^{1/\alpha}$). The time horizon considered is $T = 1$, so that $E = [0,1]$ in (8.1), and the black bars are the summands of the PSR of $\ell(T)$. Thus, when we consider the PSR of pure Lévy processes, the term $f(V_j) = \mathbb{1}(V_j \in [0,t])$ has the role of a *thinning* operator, a consideration that will be useful for the final considerations in Chapter 9.

It is clear that, while the PSR for RVs allows for a conditionally Gaussian representation of the stable distribution, the one for stochastic integrals does not. Thus a modified Poisson series representation (MPSR), that meets this requirement, was presented in Lemke and Godsill [2011] and proved in [Lemke, 2014, Chpater 4] and Lemke and Godsill [2015]. We do not report the proof here, but it is intuitive that such a representation should hold, given that, for a fixed function $\mathbf{f}$, the stochastic integral is either a stable RV or random vector, and a PSR similar to (2.29) is available for stable random vectors, see e.g. Bentkus et al. [2001].

8.1.2 MPSR for Stable Stochastic Integrals

We formulate the MPSR for

$$\tilde{\mathbf{I}}(\mathbf{f}) := \tilde{\tau}\mathbf{I}(\mathbf{f}), \tag{8.2}$$

where $\mathbf{I}(\mathbf{f})$ is defined as in (2.17)-(2.18) and $\tilde{\tau}$ is a multiplicative constant defined below; of course, the MPSR for $\mathbf{I}(\mathbf{f})$ is simply the MPSR for $\tilde{\mathbf{I}}(\mathbf{f})$ rescaled.[1] Moreover, from now on, we consider that $\alpha \neq 1$, to simplify the notation.

Let $\{\Gamma_j\}_{j=1}^{\infty}$ be, as usual, a sequence of arrival times of a unite rate Poisson process and let $\{V_j\}_{j=1}^{\infty}$ and $\{W_j\}_{j=1}^{\infty}$ be sequences of i.i.d. random variables, satisfying

$$\begin{aligned} V_j \sim \hat{m} &= \frac{m}{m(E)}, \\ \mathbb{E}[|W_j|^{\alpha}|V_j] &= \mathbb{E}[|W_j|^{\alpha}] < \infty, \\ \beta(V_j) &= \frac{\mathbb{E}[|W_j|^{\alpha}\operatorname{sgn}(W_j)|V_j]}{\mathbb{E}[|W_j|^{\alpha}]}. \end{aligned}$$

Let also

$$\tilde{\tau} := \mathbb{E}[|W_1|^{\alpha}]/C_{\alpha}, \tag{8.3}$$

with C_{α} as in (2.35). Then, the following equality in distribution holds for the SI in (8.2): $\tilde{\mathbf{I}}(\mathbf{f}) \stackrel{\mathcal{D}}{=} \mathbf{S}_2(\mathbf{f})$, where $\mathbf{S}_2(\mathbf{f})$ is the a.s. convergent random series

$$\mathbf{S}_2(\mathbf{f}) := (m(E))^{1/\alpha} \sum_{j=1}^{\infty} \left[W_j \Gamma_j^{-1/\alpha} \mathbf{f}(V_j) - b_j^{(\alpha)} \mathbb{E}[W_1]\mathbb{E}[\mathbf{f}(V_1)] \right], \tag{8.4}$$

with b_j as in (2.32).

[1] This corresponds to a rescaling of the corresponding random measure: $\tilde{\mathbf{I}}(\mathbf{f})$ is a SI computed wrt $\tilde{M} := \tilde{\tau}M$, a random measure with control measure $\tilde{m} := \tilde{\tau}m$.

From now on, we will consider m to be the Lebesgue measure, $\beta(u) = \beta$ to be constant, and $E = (0, \delta)$, as in the SI at the solution of the SDE (2.16), so that $m(E) = \delta$ and $\{V_j\} \overset{\text{i.i.d.}}{\sim} \mathcal{U}(0, \delta)$. It is clear that, if $\{W_j\} \overset{i.i.d}{\sim} \mathcal{N}(\mu_W, \sigma_W^2)$, the MPSR implies that a MaSMiN LV model holds for the stochastic integral $\tilde{\mathbf{I}}(\mathbf{f})$. In fact, conditionally on the full sequences of latent variables $\{\Gamma_j, V_j\}_{j=1}^{\infty}$, $\tilde{\mathbf{I}}(\mathbf{f})$ has Gaussian distribution

$$\tilde{\mathbf{I}}(\mathbf{f})|\{V_j, \Gamma_j\}_{j=1}^{\infty} \sim \mathcal{N}\left(\mu_W \mathbf{m}, \sigma_W^2 \mathbf{S}\right), \tag{8.5}$$

with mean vector and variance-covariance matrix proportional to the following series

$$\mathbf{m} := \delta^{1/\alpha} \sum_{j=1}^{\infty} \Gamma_j^{-1/\alpha} \mathbf{f}_j - b_j^{(\alpha)} \mathbb{E}\left[\mathbf{f}_j\right], \tag{8.6}$$

$$\mathbf{S} := \delta^{2/\alpha} \sum_{j=1}^{\infty} \Gamma_j^{-2/\alpha} \mathbf{f}_j \mathbf{f}_j', \tag{8.7}$$

in which we have defined

$$\mathbf{f}_j := \mathbf{f}(V_j), \tag{8.8}$$

a notation that we use in the rest of the chapter. As for the series representation for RVs, the infinite sequences of latent variables $\{\Gamma_j, V_j\}$ cannot be generated, meaning that the series appearing in the MPSR need to be truncated in practice.

In Lemke [2014] and Lemke and Godsill [2015] two Gaussian approximations of the residuals of the series are proposed, when f is a scalar function, in order to provide an overall approximate conditionally Gaussian structure to $\tilde{I}(f)$. The GAA scheme approximates the residual of (8.4) with a Gaussian RV, while the GAMA scheme approximates the vector of residuals of (8.6)-(8.7) with a Gaussian random vector. Our contribution is to provide the multidimensional extension of these results, as follows.

8.2 Approximation of the MPSR Residuals

As for the PSR for RVs, the initial summands of the MPSR are the larger in absolute value, and this is a first informal motivation for approximating the residual. We thus recover a similar notation to that in section 5.1, assuming that the first $N_{(0,c)}$ terms of the series are given, where $N_{(0,c)} \sim \text{Poisson}(c)$ is the random number of terms such that $\Gamma_{N_{(0,c)}} \leq c$ and $\Gamma_{N_{(0,c)}+1} > c$, and c is a truncation parameter, as before.

8.2.1 GAA for Multivariate Integrals

In analogy with the GAA for RVs presented in Section 5.1, in the GAA for SIs we spit the MPSR (8.4) into

$$\tilde{\mathbf{I}} \stackrel{\mathcal{D}}{=} \delta^{1/\alpha} \left(\sum_{j=1}^{N_{(0,c)}} W_j \Gamma_j^{-1/\alpha} \mathbf{f}_j + \mathbf{r}_{(c,\infty)} \right), \tag{8.9}$$

where the residual of the series is defined as

$$\mathbf{r}_{(c,\infty)} := \lim_{d \to \infty} \mathbf{r}_{(c,d)},$$

and[2]

$$\mathbf{r}_{(c,d)} := \sum_{j:\Gamma_j < c} W_j \Gamma_j^{-1/\alpha} \mathbf{f}_j - \sum_{j=1}^{d} b_j^{(\alpha)} \mathbb{E}[W_1] \mathbb{E}[\mathbf{f}_1]. \tag{8.10}$$

Then, denoting with $\mathrm{Var}[\cdot]$ the variance-covariance matrix,[3] in the GAA scheme $\mathbf{r}_{(c,\infty)}$ is approximated with a suitable Gaussian random vector,

$$\mathbf{r}_{(c,\infty)} \overset{\text{approx}}{\sim} \mathcal{N}(\mathbf{m}_{(c,\infty)}, \mathbf{S}_{(c,\infty)}), \tag{8.11}$$

whose mean and variance-covariance matrix are computed as[4]

$$\mathbf{m}_{(c,\infty)} := \lim_{d \to \infty} \mathbb{E}\left[\mathbf{r}_{(c,d)}\right], \quad \mathbf{S}_{(c,\infty)} := \lim_{d \to \infty} \mathrm{Var}\left[\mathbf{r}_{(c,d)}\right], \tag{8.12}$$

and whose expressions are given as follows:

Proposition 3. Let $\mathbf{m}_{(c,\infty)}$ and $\mathbf{S}_{(c,\infty)}$ be the limits in (8.12), and let us define $\mathbf{q}(\delta) := \mathbb{E}[\mathbf{f}_1]$, $\mathbf{Q}(\delta) := \mathbb{E}[\mathbf{f}_1 \mathbf{f}_1']$, with $\mathbf{f}_1$ as in (8.8). Then

$$\mathbf{m}_{(c,\infty)} = \mathbb{E}[W_1] \frac{\alpha}{1-\alpha} c^{\frac{\alpha-1}{\alpha}} \mathbf{q}(\delta),$$
$$\mathbf{S}_{(c,\infty)} = \mathbb{E}[W_1^2] \frac{\alpha}{2-\alpha} c^{\frac{\alpha-2}{\alpha}} \mathbf{Q}(\delta).$$

[2] We again assume that d is integer, so that $\lfloor d \rfloor = d$, to simplify the notation.

[3] Observe that we use the same symbol for the variance of a scalar RV and the variance-covariance matrix of a random vector.

[4] Similar to GAA for RVs, we do not claim that these are the mean and variance of $\mathbf{r}_{(c,\infty)}$. However, these quantities are likely to be the 'correct' centring and squared scaling constants for an extension of the CLT theorem 5.2.1 from Chapter 5.

Proof: We first compute the expressions of the mean and the variance-covariance matrix of $\mathbf{r}_{(c,d)}$. The mean vector is

$$\begin{aligned}\mathbb{E}[\mathbf{r}_{(c,d)}] &= \mathbb{E}\left[\sum_{j:\Gamma_j<c} W_j\Gamma_j^{-1/\alpha}\mathbf{f}_j - \sum_{j=1}^{\lfloor d\rfloor} b_j^{(\alpha)}\mathbb{E}[W_1]\mathbb{E}[\mathbf{f}_1]\right] \\ &= \mathbb{E}\left[\sum_{j=1}^{N_{(c,d)}} \mathbb{E}\left[W_jU_j^{-1/\alpha}\mathbf{f}_j\middle|N_{(c,d)}\right]\right] - \frac{\alpha}{\alpha-1}d^{\frac{\alpha-1}{\alpha}}\mathbb{E}[W_1]\mathbf{q}(\delta)\mathbb{1}(1<\alpha<2) \\ &= \mathbb{E}[N_{(c,d)}]\mathbb{E}[W_1]\mathbb{E}[U_1^{-1/\alpha}]\mathbf{q}(\delta) - \frac{\alpha}{\alpha-1}d^{\frac{\alpha-1}{\alpha}}\mathbb{E}[W_1]\mathbf{q}(\delta)\mathbb{1}(1<\alpha<2) \\ &= \mathbb{E}[W_1]\frac{\alpha}{\alpha-1}\left(d^{\frac{\alpha-1}{\alpha}} - c^{\frac{\alpha-1}{\alpha}}\right)\mathbf{q}(\delta) - \frac{\alpha}{\alpha-1}d^{\frac{\alpha-1}{\alpha}}\mathbb{E}[W_1]\mathbf{q}(\delta)\mathbb{1}(1<\alpha<2),\end{aligned} \tag{8.13}$$

where in the second equality we have used the law of total expectation, replacing the variables $\{\Gamma_j\}$ with $\{U_j\}\sim\mathcal{U}(c,d)$, after conditioning on $N_{(c,d)}$, in analogy with (5.5) - (5.7), and we have used the telescoping sum of the coefficients $\{b_j^{(\alpha)}\}$ (2.33); in the last equality we have used (5.15) and the fact that $\mathbb{E}[N_{(c,d)}] = d-c$. Taking the limit of the above expression as $d\to\infty$ yields the result claimed for $\mathbf{m}_{(c,\infty)}$.

On the other hand, given that the centring constants in (8.10) are not random,[5]

$$\mathrm{Var}[\mathbf{r}_{c,d}] = \mathrm{Var}\left[\sum_{j:\Gamma_j<c} W_j\Gamma_j^{-1/\alpha}\mathbf{f}_j\right] := \mathrm{Var}[\tilde{\mathbf{r}}_{c,d}] = \mathbb{E}[\tilde{\mathbf{r}}_{c,d}\tilde{\mathbf{r}}'_{c,d}] - \mathbb{E}[\tilde{\mathbf{r}}_{c,d}]\mathbb{E}[\tilde{\mathbf{r}}_{c,d}]',$$

where we have defined $\tilde{\mathbf{r}}_{c,d} := \sum_{j:\Gamma_j<c} W_j\Gamma_j^{-1/\alpha}\mathbf{f}_j$ to be the random part of $\mathbf{r}_{(c,d)}$. Using again the law of total expectation

$$\begin{aligned}\mathbb{E}[\tilde{\mathbf{r}}_{(c,d)}\tilde{\mathbf{r}}'_{(c,d)}] &= \mathbb{E}\left[\mathbb{E}\left[\left(\sum_{j=1}^{N_{(c,d)}} W_jU_j^{-1/\alpha}\mathbf{f}_j\right)\left(\sum_{k=1}^{N_{(c,d)}} W_kU_k^{-1/\alpha}\mathbf{f}_k\right)'\middle|N_{(0,c)}\right]\right] \\ &= \mathbb{E}\left[\sum_{j=1}^{N_{(c,d)}}\sum_{k=1}^{N_{(c,d)}} \mathbb{E}\left[W_jW_kU_j^{-1/\alpha}U_k^{-1/\alpha}\mathbf{f}_j\mathbf{f}'_k\middle|N_{(0,c)}\right]\right] \\ &= \mathbb{E}[N_{c,d}]\mathbb{E}\left[W_1^2U_1^{-2/\alpha}\mathbf{f}_1\mathbf{f}'_1\right] + \mathbb{E}[N_{c,d}^2 - N_{c,d}]\mathbb{E}\left[W_1U_1^{-1/\alpha}\mathbf{f}_1\right]\mathbb{E}\left[W_1U_1^{-1/\alpha}\mathbf{f}_1\right]' \\ &= (d-c)\mathbb{E}\left[W_1^2U_1^{-2/\alpha}\mathbf{f}_1\mathbf{f}'_1\right] + (d-c)^2\mathbb{E}\left[W_1U_1^{-1/\alpha}\mathbf{f}_1\right]\mathbb{E}\left[W_1U_1^{-1/\alpha}\mathbf{f}_1\right]' \\ &= (d-c)\mathbb{E}\left[W_1^2U_1^{-2/\alpha}\mathbf{f}_1\mathbf{f}'_1\right] + \mathbb{E}[\tilde{\mathbf{r}}_{c,d}]\mathbb{E}[\tilde{\mathbf{r}}_{c,d}]',\end{aligned}$$

[5]Recall that, in our notation, the vectors are intended as column vectors.

where the last equality follows from (8.13) and the definition of $\tilde{\mathbf{r}}_{(c,d)}$. Thus

$$\begin{aligned}\mathrm{Var}[\tilde{\mathbf{r}}_{(c,d)}] &= (d-c)\mathbb{E}\left[W_1^2 U_1^{-2/\alpha}\mathbf{f}_1\mathbf{f}_1'\right] \\ &= \mathbb{E}[W_1^2]\frac{\alpha}{\alpha-2}\left(d^{\frac{\alpha-2}{\alpha}} - c^{\frac{\alpha-2}{\alpha}}\right)\mathbf{Q}(\delta),\end{aligned} \tag{8.14}$$

and taking its limit as $d \to \infty$ leads to the statement. ■

Thus, from the GAA scheme for the residual of the MPSR for stochastic integrals it follows that, if $W_1 \sim \mathcal{N}(\mu_W, \sigma_W^2)$, we can write the approximate conditionally Gaussian representation for $\tilde{\mathbf{I}}$ (8.2). In fact, combining (8.9) and (8.11) we have:

$$\tilde{\mathbf{I}}|\{\Gamma_j, V_j\}_{j:\Gamma_j<c} \overset{\text{approx}}{\sim} \mathcal{N}\left(\delta^{1/\alpha}\left(\mu_W\mathbf{m}_{(0,c)} + \mathbf{m}_{(c,\infty)}\right),\ \delta^{2/\alpha}\left((\mu_W^2+\sigma_W^2)\mathbf{S}_{(0,c)} + \mathbf{S}_{(c,\infty)}\right)\right), \tag{8.15}$$

where we have defined

$$\mathbf{m}_{(0,c)} := \sum_{j:\Gamma_j<c} \Gamma_j^{-1/\alpha}\mathbf{f}_j, \tag{8.16}$$

$$\mathbf{S}_{(0,c)} := \sum_{j:\Gamma_j<c} \Gamma_j^{-2/\alpha}\mathbf{f}_j\mathbf{f}_j', \tag{8.17}$$

and $\mathbf{m}_{(c,\infty)}$ and $\mathbf{S}_{(c,\infty)}$ have expressions as in Proposition 3.

8.2.2 GAMA for Multivariate Integrals

As explained in Chapters 2 and 3, the GAMA scheme also allows to approximate the conditional distribution of the SI as Gaussian. We have not focussed on this approach in the rest of the thesis, for the sake of clarity. However, given that the GAMA scheme is used in Lemke [2014] and Lemke and Godsill [2015] for the implementing a Rao-Blackwellized particle filter for state inference in partially observed linear models driven by Lévy processes, in the following we give its formulation for multivariate SIs. We recall that one of the motivations behind this approach is that it allows to account for the correlation existing between (8.6) and (8.7), caused by the fact that the same random variables are involved. Moreover, this scheme also preserves the structure of (8.5), including only the parameter σ_W^2 in the variance-covariance matrix of the approximation of $\tilde{\mathbf{I}}$, as shown below, and this proves to be useful for Bayesian inference methods.

In the GAMA scheme, the joint distribution of the moments of the MPSR residual, expressed in the series (8.6) and (8.7), is modelled as multivariate Gaussian. In

particular we split these two random series as

$$\mathbf{m} = \delta^{1/\alpha}\left[\mathbf{m}_{(0,c)} + \mathbf{r}^m_{(c,\infty)}\right], \tag{8.18}$$

$$\mathbf{S} = \delta^{2/\alpha}\left[\mathbf{S}_{(0,c)} + \mathbf{R}^s_{(c,\infty)}\right], \tag{8.19}$$

with $\mathbf{m}_{(0,c)}$ and $\mathbf{S}_{(0,c)}$ as in (8.16) and (8.17), and

$$\mathbf{r}^m_{(c,\infty)} := \lim_{d\to\infty} \mathbf{r}^m_{(c,d)}, \qquad \mathbf{R}^s_{(c,\infty)} := \lim_{d\to\infty} \mathbf{R}^s_{(c,d)},$$

where $\mathbf{r}^m_{(c,d)}$ is a d-dimensional vector and $\mathbf{R}^s_{(c,d)}$ is a $d \times d$ symmetric positive definite matrix, defined as

$$\mathbf{r}^m_{(c,d)} := \sum_{j:\Gamma_j\in(c,d)} \Gamma_j^{-1/\alpha}\mathbf{f}_j - \sum_{j=1}^{d} b_j^{(\alpha)}\mathbb{E}[\mathbf{f}_j],$$

$$\mathbf{R}^s_{(c,d)} := \sum_{j:\Gamma_j\in(c,d)} \Gamma_j^{-2/\alpha}\mathbf{f}_j\mathbf{f}_j'.$$

In our computation, we vectorize the lower diagonal part of $\mathbf{R}^s_{(c,\infty)}$ and $\mathbf{R}^s_{(c,d)}$ to the $(d^2+d)/2$-dimensional[6] vectors $\mathbf{r}^S_{(c,\infty)} := \text{vech}(\mathbf{R}^S_{(c,\infty)})$ and $\mathbf{r}^S_{(c,d)} := \text{vech}(\mathbf{R}^S_{(c,d)})$, where $\text{vech}(\cdot)$ indicates the half-vectorization of a symmetric matrix.

Then the GAMA scheme aims at approximating the multivariate vector containing the residuals of the mean and variance-covariance series

$$\mathbf{r}^{\text{tot}}_{(c,\infty)} := \left[(\mathbf{r}^m_{(c,\infty)})', (\mathbf{r}^S_{(c,\infty)})'\right]' = \lim_{d\to\infty} \mathbf{r}^{\text{tot}}_{(c,d)} := \left[(\mathbf{r}^m_{(c,d)})', (\mathbf{r}^S_{(c,d)})'\right]',$$

with a 'suitable' $d + (d^2+d)/2$ Gaussian vector[7]

$$\mathbf{r}^{\text{tot}}_{(c,\infty)} \overset{\text{approx}}{\sim} \hat{\mathbf{r}}^{\text{tot}}_{(c,\infty)} := \left[(\hat{\mathbf{r}}^m_{(c,\infty)})', (\hat{\mathbf{r}}^S_{(c,\infty)})'\right]' \sim \mathcal{N}(\mathbf{m}^{\text{tot}}_{(c,\infty)}, \mathbf{S}^{\text{tot}}_{(c,\infty)}), \tag{8.20}$$

where, with obvious notation, $\hat{\mathbf{r}}^m_{(c,\infty)}$ and $\hat{\mathbf{r}}^S_{(c,\infty)}$ denote the parts of the vector approximating $\mathbf{r}^m_{(c,\infty)}$ and $\mathbf{r}^S_{(c,\infty)}$, respectively. The mean and variance-covariance of $\hat{\mathbf{r}}^{\text{tot}}_{(c,\infty)}$

[6] That is, we align in a vector the sub-diagonal columns. The resulting vector has dimension

$$d + (d-1) + \ldots + 1 = \sum_{k=0}^{d-1}(d-k) = d^2 - \left(\frac{d(d+1)}{2} - d\right) = \frac{d^2+d}{2}.$$

[7] As before, we do not claim that $\mathbf{m}^{\text{tot}}_{(c,\infty)}$ and $\mathbf{S}^{\text{tot}}_{(c,\infty)}$ are the mean and the variance-covariance matrix of $\mathbf{r}^{\text{tot}}_{(c,\infty)}$. However, these parameters are likely to be the suitable centring and scaling parameters for standardizing $\mathbf{r}^{\text{tot}}_{(c,\infty)}$ in a CLT-like theorem, whose proof is left to future studies.

are computed as[8]

$$\mathbf{m}^{\text{tot}}_{(c,\infty)} := \lim_{d\to\infty} \mathbb{E}[\mathbf{r}^{\text{tot}}_{(c,d)}] = \lim_{d\to\infty} \begin{bmatrix} \mathbb{E}[\mathbf{r}^{m}_{(c,d)}] \\ \mathbb{E}[\mathbf{r}^{S}_{(c,d)}] \end{bmatrix}, \tag{8.21}$$

$$\mathbf{S}^{\text{tot}}_{(c,d)} := \lim_{d\to\infty} \text{Var}[\mathbf{r}^{\text{tot}}_{(c,d)}] = \lim_{d\to\infty} \left[\begin{array}{c|c} \text{Var}[\mathbf{r}^{m}_{(c,d)}] & \text{Cov}[\mathbf{r}^{m}_{(c,d)}, \mathbf{r}^{S}_{(c,d)}] \\ \hline \text{Cov}[\mathbf{r}^{S}_{c,d)}, \mathbf{r}^{m}_{(c,d)}] & \text{Var}[\mathbf{r}^{S}_{(c,d)}] \end{array}\right], \tag{8.22}$$

where $\text{Cov}[\cdot,\cdot]$ is the cross-covariance matrix.[9] Their expressions are provided in the following.

Proposition 4. Let $\mathbf{m}^{\text{tot}}_{(c,\infty)}$ and $\mathbf{S}^{\text{tot}}_{(c,\infty)}$ be the limits in (8.21)-(8.22), $\mathbf{q}(\delta)$ and $\mathbf{Q}(\delta)$ as in Proposition 3, and let us define $\mathbf{g}_j := \text{vech}\left(\mathbf{f}_j\mathbf{f}_j'\right)$. Then

$$\mathbf{m}^{\text{tot}}_{(c,\infty)} = \begin{bmatrix} \frac{\alpha}{1-\alpha} c^{\frac{\alpha-1}{\alpha}} \mathbf{q}(\delta) \\ \frac{\alpha}{2-\alpha} c^{\frac{\alpha-2}{\alpha}} \text{vech}\left(\mathbf{Q}(\delta)\right) \end{bmatrix},$$

$$\mathbf{S}^{\text{tot}}_{(c,\infty)} = \left[\begin{array}{c|c} \frac{\alpha}{2-\alpha} c^{\frac{\alpha-2}{\alpha}} \mathbf{Q}(\delta) & \frac{\alpha}{3-\alpha} c^{\frac{\alpha-3}{\alpha}} \mathbb{E}\left[\mathbf{f}_1\mathbf{g}_1'\right] \\ \hline \frac{\alpha}{3-\alpha} c^{\frac{\alpha-3}{\alpha}} \mathbb{E}\left[\mathbf{g}_1\mathbf{f}_1'\right] & \frac{\alpha}{4-\alpha} c^{\frac{\alpha-4}{\alpha}} \mathbb{E}\left[\mathbf{g}_1\mathbf{g}_1'\right] \end{array}\right].$$

Proof: In the proof, the variables U_j are i.i.d. and uniformly distributed on (c,d).

Given that $\mathbb{E}[\mathbf{r}^{m}_{(c,d)}] = \mathbb{E}[\mathbf{r}_{(c,d)}]/\mathbb{E}[W_1]$, the expression of the first part of $\mathbf{m}^{\text{tot}}_{(c,\infty)}$ has already been proved in Proposition 3.

Moreover, given that the mean is an element-by-element operator, we have

$$\mathbb{E}[\mathbf{r}^{S}_{(c,d)}] = \text{vech}[\mathbb{E}(\mathbf{R}^{S}_{(c,d)})] = \text{vech}\left[\mathbb{E}\left(\sum_{j:\Gamma_j\in(c,d)} \Gamma_j^{-2/\alpha}\mathbf{f}_j\mathbf{f}_j'\right)\right] = \text{vech}\left[(d-c)\mathbb{E}\left[U_1^{-2/\alpha}\mathbf{f}_1\mathbf{f}_1'\right]\right],$$

where the last identity follows from the proof of $\mathbb{E}[\tilde{\mathbf{r}}_{(c,d)}\tilde{\mathbf{r}}_{(c,d)}]$ in Proposition 3. Taking the limit as $d\to\infty$ concludes the proof of the statement on $\mathbf{m}^{\text{tot}}_{(c,\infty)}$.

The expression of the top-left block of $\mathbf{S}^{\text{tot}}_{(c,\infty)}$ also follows from Proposition 3, by observing that $\text{Var}[\mathbf{r}^{m}_{(c,d)}] = \text{Var}[\mathbf{r}_{(c,d)}]/\mathbb{E}[W_1^2]$.

[8] The blocks of $\mathbf{S}^{\text{tot}}_{(c,\infty)}$ have dimension:

$$\left[\begin{array}{c|c} d\times d & d\times(d^2+d)/2 \\ \hline (d^2+d)/2\times d & (d^2+d)/2\times(d^2+d)/2 \end{array}\right]$$

[9] Observe that $\text{Cov}(\cdot,\cdot)$ was used before in the thesis to denote the covariance between two RVs.

The expression of the bottom-right block of $\mathbf{S}^{\text{tot}}_{(c,\infty)}$, before taking the limit, is computed as

$$\begin{aligned}
\text{Var}[\mathbf{r}^S_{(c,d)}] &= \text{Var}[\text{vech}(\mathbf{R}^S_{(c,d)})] \\
&= \text{Var}\left[\sum_{j:\Gamma_j\in(c,d)} \text{vech}(\Gamma_j^{-2/\alpha}\mathbf{f}_j\mathbf{f}_j')\right] \\
&= \text{Var}\left[\sum_{j:\Gamma_j\in(c,d)} \Gamma_j^{-2/\alpha}\mathbf{g}_j\right] \\
&= (d-c)\mathbb{E}\left[U_1^{-4/\alpha}\mathbf{g}_1\mathbf{g}_1'\right] \\
&= \frac{\alpha}{\alpha-4}\left(d^{\frac{\alpha-4}{\alpha}} - c^{\frac{\alpha-4}{\alpha}}\right)\mathbb{E}[\mathbf{g}_1\mathbf{g}_1'],
\end{aligned}$$

where we recall that $\mathbf{g}_j := \text{vech}(\mathbf{f}_j\mathbf{f}_j')$, and in the fourth equality we have adopted the same procedure as in Proposition 3 for the proof of $\text{Var}[\mathbf{r}_{(c,d)}]$. Taking the limit as $d \to \infty$ yields the result.

Finally, using the law of total expectation, the extra diagonal block of $\mathbf{S}^{\text{tot}}_{(c,\infty)}$ can be computed as the limit of

$$\begin{aligned}
\text{Cov}[\mathbf{r}^m_{(c,d)}, \mathbf{r}^S_{(c,d)}] &= \text{Cov}\left[\sum_{j:\Gamma_j\in(c,d)} \Gamma_j^{-1/\alpha}\mathbf{f}_j, \sum_{j:\Gamma_j\in(c,d)} \Gamma_j^{-2/\alpha}\text{vech}(\mathbf{f}_j\mathbf{f}_j')\right] \\
&= \mathbb{E}\left[(\tilde{\mathbf{r}}^m_{(c,d)} - \mathbb{E}[\tilde{\mathbf{r}}^m_{(c,d)}])(\mathbf{r}^S_{(c,d)} - \mathbb{E}[\mathbf{r}^S_{(c,d)}])'\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[(\tilde{\mathbf{r}}^m_{(c,d)} - \mathbb{E}[\tilde{\mathbf{r}}^m_{(c,d)}])(\mathbf{r}^S_{(c,d)} - \mathbb{E}[\mathbf{r}^S_{(c,d)}])'|N_{(c,d)}\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\tilde{\mathbf{r}}^m_{(c,d)}(\mathbf{r}^S_{(c,d)})'|N_{(c,d)}\right] - \mathbb{E}[\tilde{\mathbf{r}}^m_{(c,d)}|N_{(c,d)}]\mathbb{E}[(\mathbf{r}^S_{(c,d)})'|N_{(c,d)}]\right], \quad (8.23)
\end{aligned}$$

where we have defined $\tilde{\mathbf{r}}^m_{(c,d)} := \sum_{j:\Gamma_j \in (c,d)} \Gamma_j^{-1/\alpha} \mathbf{f}_j$, because we can neglect the constant term in $\mathbf{r}^m_{(c,d)}$ when computing the covariance. The first of these expectations is

$$\begin{aligned}
\mathbb{E}\left[\tilde{\mathbf{r}}^m_{(c,d)}(\mathbf{r}^S_{(c,d)})' | N_{(c,d)}\right] &= \mathbb{E}\left[\left(\sum_{i=1}^{N_{(c,d)}} \Gamma_i^{-1/\alpha}\mathbf{f}_i\right)\left(\sum_{j=1}^{N_{(c,d)}} \Gamma_j^{-2/\alpha}\text{vech}\left(\mathbf{f}_j\mathbf{f}_j'\right)\right)' \middle| N_{(c,d)}\right] \\
&= \sum_{i=1}^{N_{(c,d)}}\sum_{j=1}^{N_{(c,d)}} \mathbb{E}\left[U_i^{-1/\alpha}U_j^{-2/\alpha}\mathbf{f}_i\mathbf{g}_j' \middle| N_{(c,d)}\right] \\
&= \sum_{i=1}^{N_{(c,d)}}\sum_{j=1}^{N_{(c,d)}} \mathbb{E}\left[U_i^{-1/\alpha}U_j^{-2/\alpha} \middle| N_{(c,d)}\right]\mathbb{E}\left[\mathbf{f}_i\mathbf{g}_j'\right] \\
&= N_{(c,d)}\mathbb{E}\left[U_1^{-3/\alpha} \middle| N_{(c,d)}\right]\mathbb{E}\left[\mathbf{f}_1\mathbf{g}_1'\right] \\
&\quad +({N_{(c,d)}}^2 - N_{(c,d)})\mathbb{E}\left[U_1^{-1/\alpha} \middle| N_{(c,d)}\right]\mathbb{E}\left[U_1^{-2/\alpha} \middle| N_{(c,d)}\right]\mathbb{E}\left[\mathbf{f}_1\right]\mathbb{E}\left[\mathbf{g}_1'\right],
\end{aligned}$$

where, in the last equality, we have used

$$\begin{aligned}
\mathbb{E}\left[U_i^{-1/\alpha}U_j^{-2/\alpha} \middle| N_{(c,d)}\right] &= \begin{cases} \mathbb{E}\left[U_1^{-3/\alpha} \middle| N_{(c,d)}\right] & \text{if } i = j, \\ \mathbb{E}\left[U_1^{-1/\alpha} \middle| N_{(c,d)}\right]\mathbb{E}\left[U_1^{-2/\alpha} \middle| N_{(c,d)}\right] & \text{if } i \neq j, \end{cases} \\
&= \begin{cases} \frac{1}{d-c}\frac{\alpha}{\alpha-3}\left(d^{\frac{\alpha-3}{\alpha}} - c^{\frac{\alpha-3}{\alpha}}\right) & \text{if } i = j, \\ \frac{1}{(d-c)^2}\frac{\alpha^2}{(\alpha-1)(\alpha-2)}\left(d^{\frac{\alpha-1}{\alpha}} - c^{\frac{\alpha-1}{\alpha}}\right)\left(d^{\frac{\alpha-2}{\alpha}} - c^{\frac{\alpha-2}{\alpha}}\right) & \text{if } i \neq j. \end{cases}
\end{aligned}$$

Then, defining $\lambda := (d - c)$

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}\left[\tilde{\mathbf{r}}^m_{(c,d)}(\mathbf{r}^S_{(c,d)})' | N_{(c,d)}\right]\right] &= \lambda\mathbb{E}\left[U_1^{-3/\alpha} \middle| N_{(c,d)}\right]\mathbb{E}\left[\mathbf{f}_1\mathbf{g}_1'\right] \\
&\quad +(\lambda^2 + \lambda - \lambda)\mathbb{E}\left[U_1^{-1/\alpha} \middle| N_{(c,d)}\right]\mathbb{E}\left[U_1^{-2/\alpha} \middle| N_{(c,d)}\right]\mathbb{E}\left[\mathbf{f}_1\right]\mathbb{E}\left[\mathbf{g}_1'\right].
\end{aligned}$$

Furthermore, the second expectation in (8.23) is

$$\mathbb{E}[\tilde{\mathbf{r}}^m_{(c,d)}|N_{(c,d)}]\mathbb{E}[(\mathbf{r}^S_{(c,d)})'|N_{(c,d)}] = N^2_{(c,d)}\mathbb{E}\left[U_1^{-1/\alpha} \middle| N_{(c,d)}\right]\mathbb{E}\left[\mathbf{f}_1\right]\mathbb{E}\left[U_1^{-2/\alpha} \middle| N_{(c,d)}\right]\mathbb{E}\left[\mathbf{g}_1'\right],$$

so that

$$\mathbb{E}\left[\mathbb{E}[\tilde{\mathbf{r}}^m_{(c,d)}|N_{(c,d)}]\mathbb{E}[(\mathbf{r}^S_{(c,d)})'|N_{(c,d)}]\right] = (\lambda^2 + \lambda)\mathbb{E}\left[U_1^{-1/\alpha} \middle| N_{(c,d)}\right]\mathbb{E}\left[\mathbf{f}_1\right]\mathbb{E}\left[U_1^{-2/\alpha} \middle| N_{(c,d)}\right]\mathbb{E}\left[\mathbf{g}_1'\right].$$

Hence, combining these results according to (8.23) we obtain

$$\begin{aligned}\mathrm{Cov}[\mathbf{r}^m_{(c,d)}, \mathbf{r}^S_{(c,d)}] &= \lambda \Bigg[\mathbb{E}\left[U_1^{-3/\alpha} \middle| N_{(c,d)}\right] \mathbb{E}\left[\mathbf{f}_1 \mathbf{g}_1'\right] \\ &\quad - \mathbb{E}\left[U_1^{-1/\alpha} \middle| N_{(c,d)}\right] \mathbb{E}\left[\mathbf{f}_1\right] \mathbb{E}\left[U_1^{-2/\alpha} \middle| N_{(c,d)}\right] \mathbb{E}\left[\mathbf{g}_1'\right] \Bigg] \\ &= (d-c) \Bigg(\frac{\alpha}{\alpha-3} \frac{1}{d-c} \left(d^{\frac{\alpha-3}{\alpha}} - c^{\frac{\alpha-3}{\alpha}} \right) \mathbb{E}\left[\mathbf{f}_1 \mathbf{g}_1'\right] \\ &\quad - \frac{\alpha^2}{(\alpha-1)(\alpha-2)} \frac{1}{(d-c)^2} \left(d^{\frac{\alpha-1}{\alpha}} - c^{\frac{\alpha-1}{\alpha}} \right) \left(d^{\frac{\alpha-2}{\alpha}} - c^{\frac{\alpha-2}{\alpha}} \right) \mathbb{E}\left[\mathbf{f}_1\right] \mathbb{E}\left[\mathbf{g}_1'\right] \Bigg).\end{aligned}$$

Taking the limit as $d \to \infty$ we have

$$\lim_{d\to\infty} \mathrm{Cov}[\mathbf{r}^m_{(c,d)}, \mathbf{r}^S_{(c,d)}] = \frac{\alpha}{3-\alpha} c^{\frac{\alpha-3}{\alpha}} \mathbb{E}\left[\mathbf{f}_1 \mathbf{g}_1'\right],$$

as claimed. ■

Thus, from the GAMA scheme for the residual of the MPSR for stochastic integrals it follows that, if $W_1 \sim \mathcal{N}(\mu_W, \sigma_W^2)$, we can write another approximate conditionally Gaussian representation for $\tilde{\mathbf{I}}$ (8.2), alternative to (8.15). In fact, combining (8.5) with (8.18)-(8.19) and (8.20), we can write:

$$\tilde{\mathbf{I}} | \{\Gamma_j, V_j\}_{j:\Gamma_j<c} \overset{\text{approx}}{\sim} \mathcal{N}\left(\delta^{1/\alpha} \mu_W \left(\mathbf{m}_{(0,c)} + \hat{\mathbf{r}}^m_{(c,\infty)} \right),\ \delta^{2/\alpha} \sigma_W^2 \left(\mathbf{S}_{(0,c)} + \hat{\mathbf{R}}^S_{(c,\infty)} \right) \right), \tag{8.24}$$

where $\mathbf{m}_{(0,c)}$ and $\mathbf{S}_{(0,c)}$ are are defined in (8.16) and (8.17), respectively, and $\hat{\mathbf{R}}^S_{(c,\infty)}$ is the symmetric matrix such that $\mathrm{vech}(\hat{\mathbf{R}}^S_{(c,\infty)}) = \hat{\mathbf{r}}^S_{(c,\infty)}$.

Remark 12. In analogy with the GAMA scheme for the residual of the PSR for scalar RVs, there is no guarantee that the matrix $\hat{\mathbf{R}}^S_{(c,\infty)}$ is positive semidefinite, a necessary condition for a variance-covariance matrix (symmetry, on the other hand, is guaranteed by construction). This is one of the drawbacks of the GAMA wrt the GAA scheme.

8.3 MPSR for Multivariate Linear Continuous-Time Stable Models

Finally, in this section we compute the Gaussian approximation of the MPSR residual for the linear model (2.16), where $\mathbf{f}_1 = \mathbf{f}(V_1) = e^{\mathbf{A}(\delta - V_1)}\mathbf{h}$, $V_1 \sim \mathcal{U}(0, \delta)$, $\mathbf{A}$ is $d \times d$-dimensional matrix, and $\mathbf{h}$ a d-dimensional vector.

In general, given a square (real or complex valued) matrix $\mathbf{B}$, the function $y(t) := e^{\mathbf{B}t}$, $t \in \mathbb{R}$, is defined through the power series

$$e^{\mathbf{B}t} = \sum_{k=0}^{\infty} \frac{(\mathbf{B}t)^k}{k!}, \tag{8.25}$$

where $\mathbf{B}^0$ is the identity matrix that has the same size as $\mathbf{B}$. The function is often referred to as the *matrix exponential* of $\mathbf{B}$, even if the latter is properly just $e^{\mathbf{B}}$. There is a number of ways of computing or approximating (8.25), according to the properties of the matrix $\mathbf{B}$ and we refer to Moler and Van Loan [2003] for a recent review.[10] In our simulations we use the Matlab `expm` routine, that is optimized wrt the structure of the matrix in the exponent of (8.25), whenever we need to evaluate a function of this form.

Here we rather focus on the expressions involved in the GAA and GAMA approximations. In particular, in order to apply Proposition 3, we need to evaluate

$$\mathbf{q}(\delta) = \mathbb{E}\left[e^{\mathbf{A}(\delta - V_1)}\mathbf{h}\right] = \frac{1}{\delta}\int_0^{\delta} e^{\mathbf{A}(\delta - u)}\mathbf{h}\, du = \frac{1}{\delta}\int_0^{\delta} e^{\mathbf{A}t}\mathbf{h}\, dt, \tag{8.26}$$

where the last equality follows by the simple change of variable $t = \delta - u$, and

$$\mathbf{Q}(\delta) = \mathbb{E}\left[e^{\mathbf{A}(\delta - V_i)}\mathbf{h}\mathbf{h}'\left(e^{\mathbf{A}(\delta - V_i)}\right)'\right] = \frac{1}{\delta}\int_0^{\delta} e^{\mathbf{A}t}\mathbf{h}\mathbf{h}'\left(e^{\mathbf{A}t}\right)' dt, \tag{8.27}$$

that are integrals of matrix exponentials on a bounded support. According to the structure of $\mathbf{A}$, these integrals can be computed in a number of ways.[11] We proceed using the procedure proposed by Van Loan [1978], that can be applied regardless of the structure of $\mathbf{A}$. This method involves computing the exponential of auxiliary matrices and combining its blocks to obtain (8.26) and (8.27). In particular, defining

[10] For example, if $\mathbf{B}$ is diagonalizable, one needs to compute only the matrix exponential of the diagonal matrix with the eigenvalues on the diagonal, which has a trivial expression.

[11] For example, if $\mathbf{A}$ is invertible, we have that $\mathbf{q}(\delta) = \frac{1}{\delta}\left[-\mathbf{A}\right]^{-1}\left[e^{\mathbf{A}t}\right]_0^{\delta}\mathbf{h} = \frac{1}{\delta}\mathbf{A}^{-1}\left[e^{\mathbf{A}\delta} - \mathbf{I}\right]\mathbf{h}$, using properties of (8.25). Moreover, given that the integral of a vector or matrix-valued function is its component-wise integral, (8.26) and (8.27) could be computed numerically.

the following $(d+d+1)$ and $(d+d)$-dimensional square matrices,

$$\mathbf{C}_1 := \begin{bmatrix} -\mathbf{A}' & \mathbf{h}\mathbf{h}' & \mathbf{0}_{d\times 1} \\ \mathbf{0}_{d\times d} & \mathbf{A} & \mathbf{h} \\ \mathbf{0}_{1\times d} & \mathbf{0}_{1\times d} & 0 \end{bmatrix}, \qquad \mathbf{C}_2 := \begin{bmatrix} \mathbf{A} & \mathbf{h}\mathbf{h}' \\ \mathbf{0}_{d\times d} & -\mathbf{A}' \end{bmatrix}, \tag{8.28}$$

where $\mathbf{0}_{i\times j}$ is a $i \times j$ null matrix, then their exponentials have block structures

$$e^{\mathbf{C}_1\delta} = \begin{bmatrix} \mathbf{F}_2(\delta) & \mathbf{G}_2(\delta) & \mathbf{H}_2(\delta) \\ \mathbf{0}_{d\times d} & \mathbf{F}_3(\delta) & \mathbf{G}_3(\delta) \\ \mathbf{0}_{1\times d} & \mathbf{0}_{1\times d} & F_4(\delta) \end{bmatrix}, \qquad e^{\mathbf{C}_2\delta} = \begin{bmatrix} \mathbf{F}_3(\delta) & \mathbf{G}_4(\delta) \\ \mathbf{0}_{d\times d} & \mathbf{F}_5(\delta) \end{bmatrix}, \tag{8.29}$$

where the blocks have expressions given in Van Loan [1978] (see also also Wahlström et al. [2014]). Specifically, it holds that

$$\mathbf{q}(\delta) = \frac{1}{\delta}\mathbf{G}_3(\delta), \tag{8.30}$$

$$\mathbf{Q}(\delta) = \frac{1}{\delta}\mathbf{G}_4(\delta)\mathbf{F}_3'(\delta), \tag{8.31}$$

$$e^{\mathbf{A}\delta} = \mathbf{F}_3(\delta). \tag{8.32}$$

In this way, all we have an estimate for the moments of the GAA scheme, by simply compute an accurate approximation of $e^{\mathbf{C}_1\delta}$ and $e^{\mathbf{C}_2\delta}$.

On the other hand, recalling that $\mathbf{g}_1 = \text{vech}(\mathbf{f}_1\mathbf{f}_1')$, Proposition 4 requires to evaluate also

$$\mathbb{E}[\mathbf{f}_1\mathbf{g}_1'] = \mathbb{E}\left[\left(e^{\mathbf{A}(\delta-V_i)}\mathbf{h}\right)\left(\text{vech}\left(e^{\mathbf{A}(\delta-V_1)}\mathbf{h}\mathbf{h}'e^{\mathbf{A}'(\delta-V_1)}\right)\right)'\right],$$

and

$$\mathbb{E}[\mathbf{g}_1\mathbf{g}_1'] = \mathbb{E}\left[\text{vech}\left(e^{\mathbf{A}(\delta-V_i)}\mathbf{h}\mathbf{h}'\left(e^{\mathbf{A}(\delta-V_i)}\right)'\right)\text{vech}\left(e^{\mathbf{A}(\delta-V_i)}\mathbf{h}\mathbf{h}'\left(e^{\mathbf{A}(\delta-V_i)}\right)'\right)'\right].$$

We were not able to find a matrix exponential approach analogous to that used for Proposition 3, so we approximated these quantities numerically by taking the element-wise integral of the corresponding matrices.[12]

[12] We remark only that the half vectorization operator was implemented using the vectorization operator and the Elimination and Duplication sparse matrices, see Magnus and Neudecker [1980].

8.3.1 Approximate Simulation from the Process

In this section we show approximate simulations from linear processes driven by stable Lévy noise, using the MPSR of the corresponding multivariate stable distribution, and both the GAA and the GAMA schemes for approximating its residual. In particular, we consider

$$\mathbf{h} = [0, \dots, 0, 1]',$$

representing the fact that the Lévy process has direct effect only on the last component of the state,[13] and three scenarios for the matrix $\mathbf{A}$:

$$\mathbf{A}_1 = \left[\begin{array}{c|c} \mathbf{0}_{(P-1)\times 1} & \mathbf{I}_{(P-1)\times(P-1)} \\ \hline -a_P & -a_{P-1} \dots - a_1 \end{array}\right], \quad \mathbf{A}_2 = \begin{bmatrix} \theta_1 & 1 \\ 0 & \theta_2 \end{bmatrix}, \quad \mathbf{A}_3 = \begin{bmatrix} 0 & 1 \\ 0 & \theta \end{bmatrix},$$

where $\mathbf{0}_{(P-1)\times 1}$ is a $(P-1)$-dimensional null vector, $\mathbf{I}_{(P-1)\times(P-1)}$ is the $(P-1)\times(P-1)$ identity matrix. When used in the linear model (2.16), with $\mathbf{x}(t) = [x_1(t), \dots, x_P(t)]'$, these transition matrices correspond to the following scenarios:

- $\mathbf{A}_1$ corresponds to a continuous-time autoregressive model of order P, CAR(P), where $x_1(t)$ is the state and $x_i(t), i > 1$, are its i-th derivatives, see Brockwell [2001, 2004]; Brockwell and Lindner [2009]. The linear dynamic system with matrix $\mathbf{A}_1$ is stable if the parameters $\{a_i\}_{i=1}^P \in \mathbb{R}^+$. The matrix $\mathbf{A}_1$ is also denoted *companion matrix* in the literature, and it is generally non invertible. However, if its eigenvalues $\lambda_1, \dots, \lambda_P$ are distinct, $\mathbf{A}_1$ is diagonalizable as $\mathbf{A}_1 = \mathbf{V}^{-1}\text{diag}(\lambda_1, \dots, \lambda_P)\mathbf{V}$, where $\mathbf{V}$ is the Vandermonde matrix corresponding to the eigenvalues, see Horn et al. [1990]; thus, in our simulations we typically fix the eigenvalues, and find the corresponding parameters a_i of $\mathbf{A}_1$. Stability is equivalently achieved if the eigenvalues have negative real part;[14]

- $\mathbf{A}_2$ corresponds to a model with both components reverting to their mean value, through the parameters θ_1 and θ_2, and a 'trend effect' on $x_1(t)$. Observe that, because of the mean-reversion parameter θ_1, $x_2(t)$ does not have the meaning of derivative of $x_1(t)$. The dynamic system with matrix $\mathbf{A}_2$ is stable if $\{\theta_i\}_{i=1}^2$, coinciding with the eigenvalues of the process, have negative real part. Such model could represent the behaviour of a financial time series, as in Christensen et al. [2012]; Murphy [2014]. Notice however that in this literature, a jump-diffusion model is assumed, that adds a jump process to the Gaussian diffusion, in order to account for extreme observations. On the contrary, in our

[13] However, the expressions that we provide are generic.
[14] See for example Franklin et al. [1994] or Aström and Murray [2010] .

approach this is simply achieved if the tail parameter α of the Lévy process is smaller than 2. We leave a more thorough comparison of the two approaches to future work;

- $\mathbf{A}_3$ corresponds to a Langevin model, see Lemons and Gythiel [1997]. In this case, $x_1(t)$ is subject only to a 'trend effect', or, equivalently, $x_2(t)$ has the meaning of derivative of $x_1(t)$, while $x_2(t)$ is reverting to its mean through the coefficient θ. The eigenvalues of the transition matrix are 0 and θ, so that the system is marginally stable.

Figure 8.2 shows approximate simulations from bi-dimensional linear models with matrices $\mathbf{A}_1$, $\mathbf{A}_2$ and $\mathbf{A}_3$, for 1000 irregularly sampled time steps, initialized at $\mathbf{x}_0 = \mathbf{0}$. The driving Lévy processes has tail parameter $\alpha = 0.8$ in the top row of sub-figures and $\alpha = 1.5$ in the bottom row, while the RVs $\{W_j\}$ in the MPSR are Gaussian with $\mu_W = 1$ and $\sigma_W = 1$. The parameters of the matrices are specified in the titles of each sub-figure. For each approximate simulation of the process, we display the MPSR simply truncated to $c = 50$, and the MPSR with the GAA and GAMA of its residual.

It is possible to represent a number of different situations by changing the transition matrix. For example, Figure 8.3 shows that complex eigenvalues in a CAR(4) system produce oscillatory states (the values in the matrix are $a_4 = 0.0096$, $a_3 = 0.1040$, $a_2 = 0.4800$ and $a_1 = 1.1000$, corresponding to the eigenvalues $\lambda_1 = -0.2 + 0.2i$, $\lambda_2 = -0.2 - 0.2i$, $\lambda_3 = -0.3$ and $\lambda_4 = -0.4$; the Lévy process has tail parameter $\alpha = 1.5$, and the MPSR is truncated at $c = 50$).

From these figures, we observe that the three realizations (truncated MPSR and truncated MPSR with either the GAA or the GAMA scheme for the residual) appear similar for smaller values of α, while it is possible to notice some differences when α is closer to 2. This is consistent with our findings on the PSR for RVs in Chapters 6 and 7. The contribution of the residuals to the full series is considerably smaller for distributions with heavier tails (small α): in fact in this case, the first parts of the series are very close to the full series, and approximating the residuals has a small effect.

On the other hand, again in analogy with our studies on scalar RVs, experimental results shows that the distribution of residuals of the series is closer to a Gaussian when α is closer to 2, as follows.

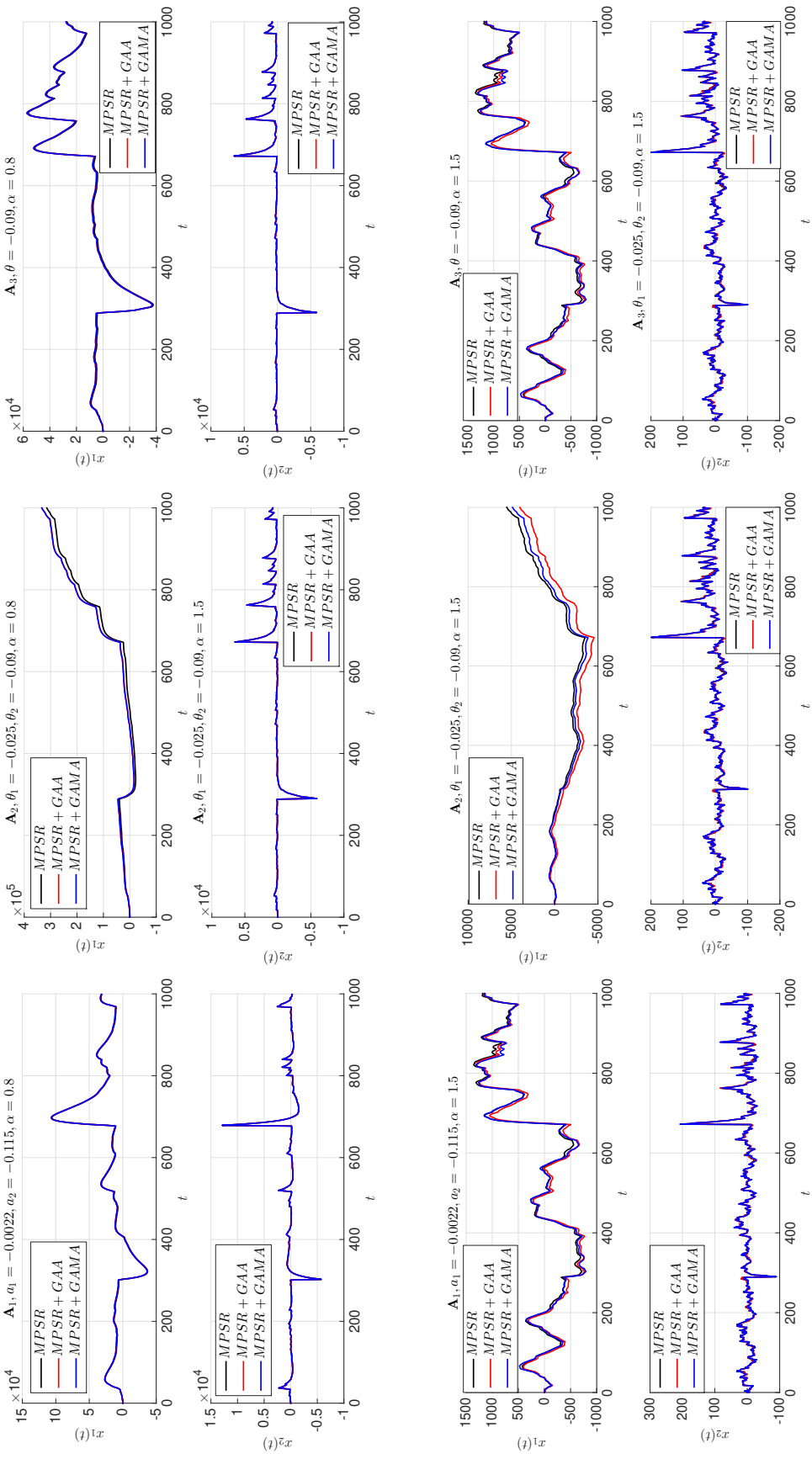


Figure 8.2 Approximate simulations of bi-dimensional stable linear systems. In each sub-figure we plot the truncated MPSR, and the MPSR with either the GAA and the GAMA scheme for the residuals. We consider processes corresponding to all three matrices $\mathbf{A}_1$, $\mathbf{A}_2$ and $\mathbf{A}_3$ (see the title of each sub-figure for the parameters), and two scenarios for the driving stable Lévy process: more heavy tailed in the top row of sub-figures, and less heavy-tailed in the bottom row of sub-figures.

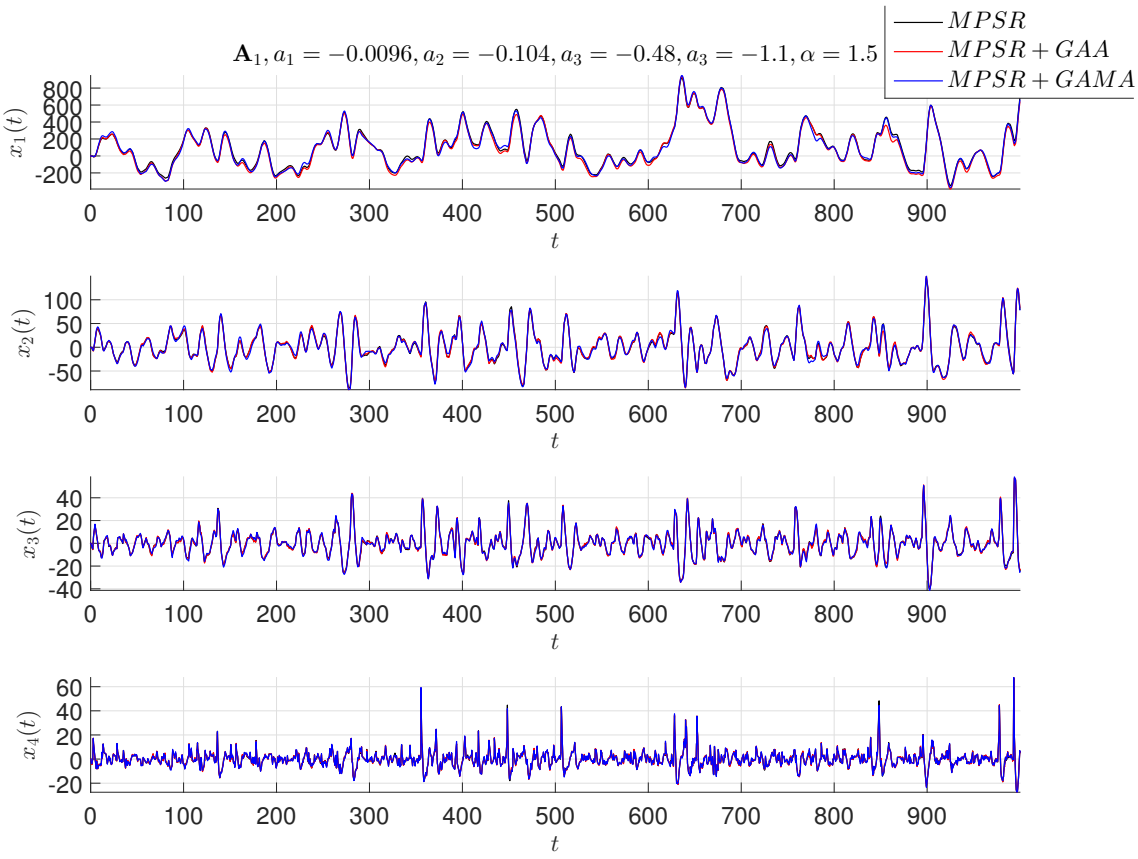


Figure 8.3 Approximate simulation of a four-dimensional stable linear system with oscillations. In each sub-figure we plot the truncated MPSR, and the MPSR with either the GAA and the GAMA scheme for the residuals. We consider a process corresponding to the matrix $\mathbf{A}_1$, with parameters a_i and complex eigenvalues as in the text. The driving stable Lévy process has tail parameter $\alpha = 1.5$.

8.3.2 Validation of the MPSR Residual Approximation

For conciseness, here we experimentally validate the GAA scheme for $\mathbf{r}_{(c,\infty)}$, as in (8.11), but analogous simulations can be done for $\mathbf{r}^{\text{tot}}_{(c,\infty)}$ in the GAMA scheme (8.20).

To test the multivariate Gaussianity of $\mathbf{r}_{(c,\infty)}$, we generate 10^3 samples of the MPSR residual between $c = 100$ and $d = c + 10^4$, $\mathbf{r}_{(c,d)} = [r^1_{(c,d)}, r^2_{(c,d)}]$, corresponding to the CAR(2) model, with eigenvalues $\lambda_1 = -0.2$, $\lambda_2 = -0.3$, or, equivalently, $a_2 = 0.06$ and $a_1 = 0.5$. The stochastic integrals are computed over the interval $(0, \delta) = (0, 1)$. Furthermore, in the MPSR we choose $W_j \sim \mathcal{N}(1, 1)$.

Figure 8.4 shows the contours of the sample distribution of the bivariate data, when the tail parameter of the driving Lévy process is $\alpha = 1.5$: the sample distribution of $\mathbf{r}_{(c,d)}$ resembles a Gaussian. This is confirmed by Figure 8.5, where we compare the sample marginal distributions of $\mathbf{r}_{(c,d)}$ and those of a bivariate Gaussian with mean $\mathbb{E}[\mathbf{r}_{(c,d)}]$ as in (8.13) and variance-covariance matrix $\text{Var}[\mathbf{r}_{(c,d)}] = \text{Var}[\tilde{\mathbf{r}}_{(c,d)}]$ as in (8.14).

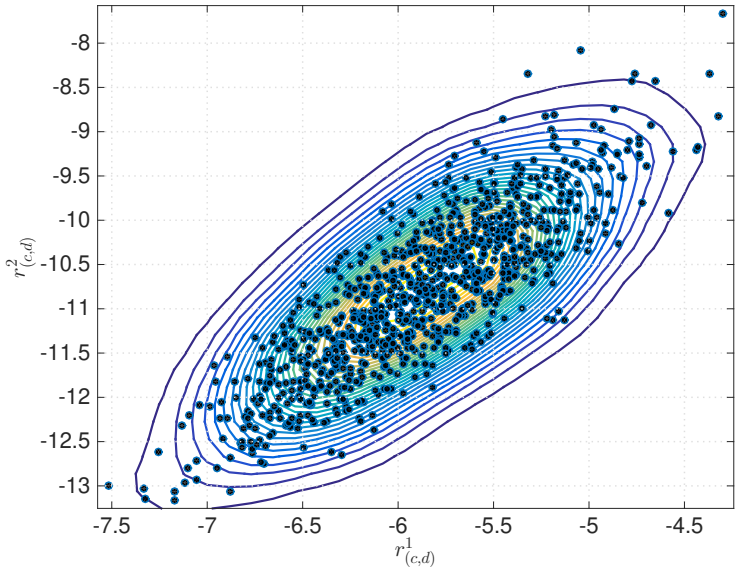


Figure 8.4 Contour plot of the sample distribution of $\mathbf{r}_{(c,d)}$, overlaid with 10^3 samples. See the text for the values of the parameters.

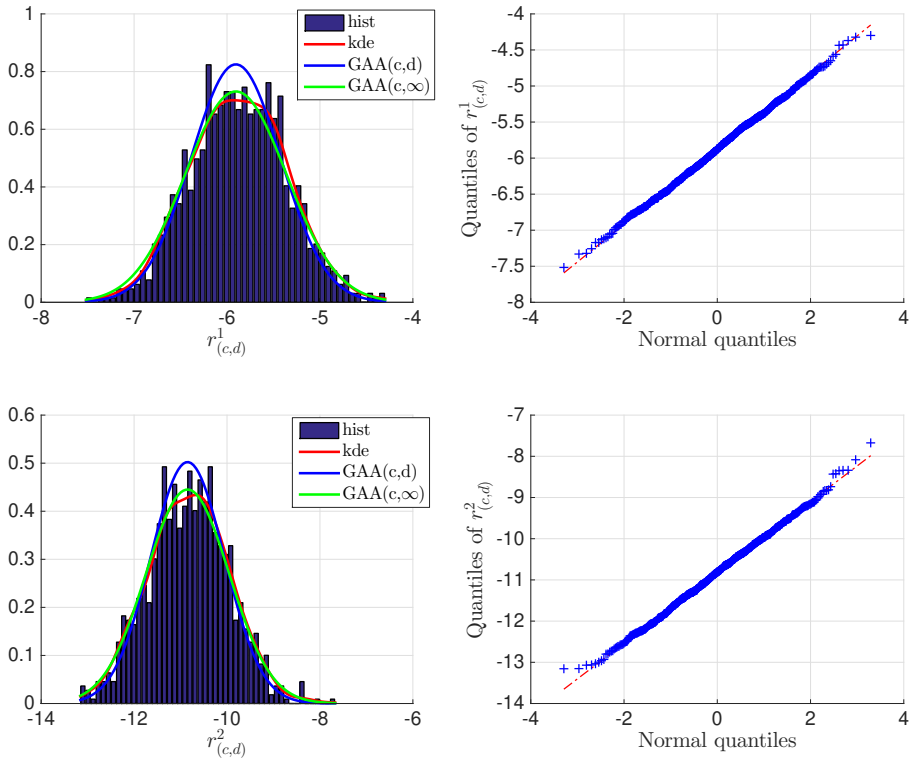


Figure 8.5 Analysis of the sample marginal distributions of $\mathbf{r}_{(c,d)}$ for the MPSR of the CAR(2) process with parameters in the text. Left: histograms and kernel density estimates (kde) overlaid to the Gaussian PDFs, with moments computed on the interval (c, d), as in (8.13) and (8.14), and on (c, ∞), as in Proposition 3. Right: comparison of the quantiles of $r^1_{(c,\infty)}$ and $r^2_{(c,\infty)}$ with those of a Gaussian with fitted moments.

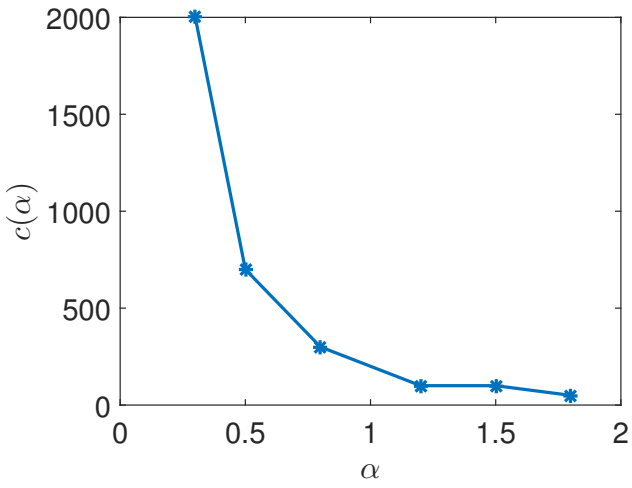


Figure 8.6 Values of c that enable multivariate normality of $\mathbf{r}_{(c,d)}$ in the Royston test, as a function of and $\alpha \in \{0.3, 0.5, 0.8, 1.2, 1.5, 1.8\}$. The parameters of $\mathbf{r}_{(c,d)}$ and the corresponding CAR(2) process are given in the the text.

For completeness, we also plot the curves corresponding to a Gaussian distribution with mean $\mathbf{m}_{(c,\infty)}$ and variance-covariance matrix $\mathbf{S}_{(c,\infty)}$, as in Proposition 3. Despite the good fits, we recall that these are only necessary, not sufficient conditions for (or, rather, indications of) multivariate Gaussianity of $\mathbf{r}_{(c,\infty)}$.

Finally, we perform the Royston's test, Royston [1983],[15] on set of samples of $\mathbf{r}_{(c,d)}$, each of size 1000, related to the MPSR of the CAR(2) linear process considered above. The test was run over a grid of values of c, for the range of values $\alpha \in \{0.3, 0.5, 0.8, 1.2, 1.5, 1.8\}$ of the tail parameter of the driving Lévy process. Figure 8.6 shows a star in correspondence of the value of c after which the Gaussianity hypothesis was almost never rejected. The results indicate that the distribution of the MPSR residual appears to be closer to a Gaussian for larger values of α, in agreement with our findings for the PSR residual in Chapter 6.

The latter is, of course, a simulation intensive and non-exact procedure. However, it encourages future research into the formulation of a CLT for the MPSR residual, as well as of systematic, analytical bounds, to be used for practical inference purposes, in analogy with those that we developed scalar stable RVs.

[15]This is a multivariate generalization of the Shapiro and Wilk's test Shapiro and Wilk [1965].

Chapter 9

Conclusions and Future Work

Stable distributions are extremely important, given their role in the generalized central limit theorem, and their ability to model heavy-tailed and skewed data. However, this class of laws is intractable, because it lacks closed form expressions for the quantities used in simple and standard inference procedures. Nonetheless, latent variable models for stable distributions are available, and they suit the framework of Bayesian inference, which is also favored by the heavy-tailed nature of the distribution. In this thesis we have provided a number of results towards inference schemes for some of these latent variable models, either targeting exact methods, or providing bounds on the error introduced by approximate methods. We conclude by summarizing our findings, pointing out the novel approaches that they enable, together with their current limitations, and providing some ideas for future research directions on the topic.

9.1 Pseudo-Marginal Parameter Sampler

Our first contribution was the formulation, in Chapter 4, of a novel inference scheme for the posterior of the scalar distribution parameters, based on the CMS LV model. We showed that the adaptive proposals-based PM method is able to achieve results comparable to the Bayesian approach presented by Buckle [1995]. This was obtained using an off-the-shelf implementation of the PM method, i.e., *without* requiring the application-specific re-parametrization of the parameter full-conditionals, that enables the mixing of the GS, but that also introduces a bias in the estimates. This constitutes a first formal advantage of the PM versus the GS scheme, even though simulation results show that the bias in the GS appears to be negligible, if the function $t_{\alpha,\beta}$ is inverted through an accurate numerical method.

Furthermore, the PM method has the advantage that the design parameters G and M directly influence the mixing of the Markov chains; larger values result in better mixing and these parameters can thus be chosen as large as possible w.r.t. the available computational resources. The mixing of the GS scheme, however, is strictly limited by the dependencies between the latent variables and the parameters. It is not possible to improve upon this unless the user is able to come up with an even "better" re-parametrization.

The latter point could prove to be an even larger merit of the pseudo-marginal method in more challenging scenarios, e.g. for models including more parameters or where the posterior distribution has significant probability mass on both disconnected subsets of the support S_α, forcing the sampler to jump between these subsets. The conditioning on the latent variables may prohibit large moves in the parameter space, in particular between disconnected subsets of the posterior support, which is not an issue for a marginal sampler employing global (i.e., independent) proposals. Investigating the performance of the methods in such more challenging scenarios is a topic for future work.

However, it is necessary to underscore the major drawback of the novel method, resulting in an increased computational cost. For each MCMC iteration, in the GS each of the N latent variables is drawn only once by rejection sampling, and each acceptance requires, on average, no more than 6 candidate points. On the other hand, $M \gg 1$ draws of each of the N latent variables are required in the PM scheme, after G adaptations of the envelope function. Thus, if we consider that the computationally expensive part of the IS is the adaptation of the proposal distribution, the cost of one iteration of the PM scheme is $G \times N$, compared to $6 \times N$ for the GS; if, on the other hand, we want to take into account the number of evaluations of the kernel of the full-conditional distribution of the LVs, then the cost of each PM iteration is $(G + M) \times N$, compared to $7 \times N$ for the GS.

If one chooses to work with the PM scheme, adaptive proposal distributions of the LVs are shown to lead to smaller variance of the likelihood estimator, compared to more standard proposals (Uniform and Laplace approximation proposals), but they have higher computational cost. Moreover, there seems to be room for improvement (and possibly optimization) of the performances by changing the adaptive scheme. We illustrated this principle by increasing the order of the proposal from piece-wise constant to piece-wise linear, but a number of alternative schemes can be easily implemented (consider e.g. a an hybrid between the Laplace approximation and adaptive proposals).

Were an equally efficient but less expensive approach for the IS step to be found, this would likely improve the state of the art of Bayesian inference of the α-stable

parameters, based on the CMS LV model (2.23). Moreover, we remark that the development of adaptive proposals was instrumental to the application of the PM method to inference for α-stable parameters, but this adaptive scheme could find applications also in other contexts.

9.2 Approximations of the Poisson Series Representation

In Chapters 5 - 7 we worked on the Poisson series representation for stable RVs, which enables the MaSMiN latent variable model, together with the inference methods that are used for conditionally Gaussian distributions.

Despite being a very appealing framework, in practice the PSR needs to be truncated, because it is not possible to generate infinite sequences of RVs, and this induces an approximation error. In Chapter 5, we formalized a CLT-like result for the residual of the series, as the truncation parameter is set to infinity. This is an asymptotic result, that justifies the Gaussian approximation approach for the PSR residual, for any distribution on W_1 such that $\mathbb{E}[|W_1|^3] < \infty$. In fact, the GAA scheme had been adopted in practice in previous literature aimed at developing approximate inference methods, and our work serves as a formal justification for it.

We then focussed on the case of Gaussian RVs $\{W_j\}$, which enables the MaSMiN model. The tool that we used to quantify the error caused by the finite truncation of the series, is the Esséen's smoothing lemma, which transfers information of convergence between characteristic functions to CDFs. For this purpose, we provided series and integral expressions for the CFs of the RVs involved in the approximation scheme ($X_{(0,c)}$ and $R_{(c,\infty)}$).[1] In particular, for the case of symmetric $\{W_j\}$, corresponding to symmetric stable laws, we were able to obtain closed-form expressions of these CFs.

These were then used to establish analytical bounds on the Kolmogorov distances of interest. First, in Chapter 6, we bounded the distance between the PSR residual from a Gaussian, obtaining two bounds, one of asymptotic order $O(1/c)$ and one of asymptotic order $O(1/\sqrt{c})$. The former is, of course, asymptotically superior, and it reflects the fact that, in this specific case, the third moment of the RVs $\{W_j\}$ is zero, allowing one to obtain a stronger rate than what could be typically expected from the Esséen's smoothing lemma. However, this bound is lose for small values of α and c, in which case, the bound of order $O(1/\sqrt{c})$ provides a sharper result.

[1] We stress that the CFs are expressed with respect to the parameters μ_W and σ_W^2, instead of σ and β, according to the re-parametrization induced by the PSR.

These bounds on the PSR residual allowed us to construct further bounds on the distance between the truncated PSR with added residual and the full PSR, by again applying the smoothing lemma in Chapter 7. Moreover, we re-visited an existing literature result on the distance between the simply truncated PSR and the stable law, adapting it to our case of random number of terms in the truncation. Such a result does not follow from the smoothing lemma, but it gives bounds with order that is asymptotic to those deriving from the smoothing lemma, as showed by numerical simulations. Finally, we established that, for a large range of the parameters c and α, the Gaussian approximation of the PSR residual produces a smaller error than simply truncating the series, suggesting that the GAA scheme should be adopted in practical inference procedures. All the established analytical results are in agreement with the numerically simulated bounds.

We remark that, while an exact scale mixture of normals representation is enabled for symmetric stable distributions, this is not the case for the asymmetric distributions, making an extension of our work to arbitrary values of μ_W and σ_W of primary relevance. Moreover, in this thesis we did not deal with the GAMA scheme for the PSR residual, and this is also a topic of future research. We expect that the current results and our future extensions will form a collection of tools that will be used for the selection of the PSR truncation parameter c, so as to control the quality of the approximation induced.

Finally, were similar analytical bounds on the distance between PDFs to be established, we envisage the possibility of developing novel exact inference methods for stable distributions, using retrospective sampling methods for iterative generation of a finite number of terms from the series. A first result towards this goal has been presented in the thesis by generating numerical simulations of a uniform distance of the PDF of the PSR residual from a Gaussian. We leave to future work the formulation of analytical, non-uniform, bounds, and the development of inference methods.

9.3 Approximate Series Representations for Multivariate Stable Processes and Distributions

Our contribution in Chapter 8 was the approximation of the MPSR for multivariate stable stochastic integrals, as an extension of the work on scalar integrals presented in recent literature. The MPSR is of great relevance for Bayesian state inference in continuous-time stochastic models driven by stable noise. In fact, a conditionally Gaussian representation of the process follows from the MPSR, and this allows to

reduce the Monte Carlo variance of SMC filters, by incorporating Kalman-filter steps for the Gaussian part of the state. However, similar to the PSR, the MPSR cannot be exactly generated.

Our first contribution was the formulation of both the GAA and the GAMA schemes for approximating the MPSR residuals, while maintaining an (approximate) conditionally Gaussian representation of the corresponding stochastic integral. This allows for approximate simulation from the process, alternative to the simple truncation of the series, which is adopted in the literature.

Our guess is that our approach reduces the distance between the distribution of the approximated process and the exact model wrt this more common approach. As a first step towards verifying this, we performed simulations to test the Gaussianity of the residuals of the MPSR of multivariate linear processes, and to analyse how this varies with α, the tail parameter of the driving Lévy process. In the future, we will formalize a CLT for the MPSR and multivariate MPSR residuals, and we will establish analytical bounds similar to those provided for the PSR of scalar stable RVs, for a systematic study of the convergence of our approximate representation of continuous-time processes.

Again in analogy to the case of RVs, all such bounds could be used to control the quality of the inference for stochastic models driven by stable Lévy noise. Comparisons, in terms of both modelling and inference, with alternative models aimed at representing extreme events, such as the jump-diffusion models, is also left as a topic for future work.

A formulation of the GAA and GAMA schemes for the residuals of the PSR for multivariate distributions is still lacking, but it could be be easily provided, in analogy with our work on the multivariate MPSR. As anticipated in Chapters 1 and 2, in this case the series representation reparametrizes the multivariate stable distribution, allowing one to avoid dealing with the infinite dimensional spectral measure. This will simplify the inference for stable vectors, a problem that is currently made tractable by either assuming that the spectral measure is discrete or that the multivariate stable distribution is sub-Gaussian. The inference schemes resulting from our approach will be, in first instance, approximate, but we envisage the possibility of controlling their quality by again establishing bounds on the error committed.

We conclude by emphasising some connections between the series representations for stochastic processes and variables. The nonasymptotic bounds that we developed for the PSR of RVs could be easily extended to the MPSR of pure Lévy processes, thanks to the considerations in Section 8.1.1: the integrand, in this case, acts as a

thinning operator, that selects only some of the summands of both the PSR and MPSR, whose structure then coincides with that of a PSR for simple RVs.

The MPSR of SIs with a generic integrand does not relate so trivially to the PSR of RVs. Nonetheless, we recall that for a fixed integrand f, scalar SIs are stable RVs, whose distribution parameters depend on f, as explained in Section 2.3.3. This enables one to use our bounds for the PSR of scalar RVs also for scalar SIs, upon considering the dependence of the parameters on f. Similarly, were bounds for the PSR of multivariate stable distributions to be developed, now also as a function of the dimension, these could be, in theory, adapted to multivariate SIs, that are stable vectors, once the integrand $\mathbf{f}$ is fixed. However, the mapping of the relative parameters is not as simple as in the scalar case, and it might be better to establish bounds related to the MPSR directly, when dealing with multivariate stable SIs.

Appendix A

Further Considerations on the LV Models for Stable Distributions

A.1 Transformations for σ and β in the MaSMiN Model

Here we consider how the mappings expressed in (2.34a) and (2.34b) allow to obtain any value of $\sigma > 0$ and $\beta \in [-1, 1]$, given the tail parameter α and the distribution of W_1, when the latter is continuous and belonging to a location-scale family, with PDF $f_W(w)$, $w \in \mathbb{R}$, as in the case examined in the paper $W_1 \sim \mathcal{N}(\mu_W, \sigma_W^2)$. We introduce the following auxiliary PDF

$$\pi(w) := \frac{|w|^\alpha f_W(w)}{\int_{\mathbb{R}} |w|^\alpha f_W(w)\, \mathrm{d}w} = \frac{\tilde{\pi}(w)}{I},$$

where $\tilde{\pi}(w)$ and I denote the unnormalized density and the normalizing constant, respectively. Then (2.34a), the transformation related to σ, can be rewritten as

$$\sigma = \frac{\int_{\mathbb{R}} \tilde{\pi}(w)\, \mathrm{d}w}{C_\alpha} = \frac{I}{C_\alpha}.$$

Given that $C_\alpha > 0$, it results that $\sigma > 0$. Moreover, for any fixed value of α (and consequently C_α), it is possible to achieve any $\sigma > 0$, by determining the parameters of the distribution of W_1 that give the necessary value of I; the reader can think, for example, to the scale parameter, corresponding to σ_W in our setting. On the other

hand, we can express (2.34b), the transformation related to β, as

$$\begin{aligned}\beta &= \frac{-\int_{-\infty}^{0}\tilde{\pi}(w)\,\mathrm{d}w + \int_{0}^{\infty}\tilde{\pi}(w)\,\mathrm{d}w}{I} \\ &= = -\int_{-\infty}^{0}\pi(w)\,\mathrm{d}w + \int_{0}^{\infty}\pi(w)\,\mathrm{d}w \\ &= -\left(1-I^{+}\right)+I^{+} \\ &= 2I^{+}-1, \end{aligned} \tag{A.1}$$

where I^{+} is the probability mass assigned by $\pi(w)$ to $\mathbb{R}^{+}$. Then any $\beta \in [-1,1]$ can be obtained by setting the parameters of the distribution of W_1 to obtain the necessary $I^{+} \in [0,1]$; the reader can think for example to the location parameter, corresponding to μ_W in the Gaussian case. In particular:

- Given that $|w|^{\alpha}$ is a symmetric function, it is clear from (2.34b) that $\beta = 0$ when $f_W(w)$ is an even function. In our Gaussian setting on W_1, this corresponds to $\mu_W = 0$.[1] Furthermore, when $\mu_W = 0$, using (2.34a) and (2.35) and the expression of the α-th absolute moment of the Gaussian central distribution Winkelbauer [2012], defining $a := \alpha/2$, and recalling that $\Gamma(\cdot)$ is the Gamma function (B.12), Appendix B.4,

$$\begin{aligned}\sigma^{\alpha} = \frac{\mathbb{E}[|W|^{\alpha}]}{C_{\alpha}} &= \sigma_W^{\alpha}\frac{2^{\alpha/2}\Gamma\left(\frac{\alpha+1}{2}\right)}{\sqrt{\pi}}\frac{\Gamma(2-\alpha)\cos(\pi\alpha/2)}{1-\alpha} \\ &= \sigma_W^{\alpha}\frac{2^{a}\Gamma\left(a+\frac{1}{2}\right)}{\sqrt{\pi}}\frac{\Gamma(2-2a)\cos(\pi a)}{1-2a} \\ &= \sigma_W^{\alpha}\frac{2^{1-a}\Gamma(2a)}{\Gamma(a)}\Gamma(1-2a)\cos(\pi a) \\ &= \sigma_W^{\alpha}\frac{2^{1-a}\pi\cos(\pi a)}{\Gamma(a)\sin(2\pi a)} \\ &= \sigma_W^{\alpha}\frac{2^{-a}\pi}{\Gamma(a)\sin(\pi a)} \\ &= \sigma_W^{\alpha}2^{-a}\Gamma(1-a) \\ &= \sigma_W^{\alpha}2^{-\alpha/2}\Gamma(1-\alpha/2), \end{aligned} \tag{A.2}$$

  where in the fourth equality we have used the duplication formula for Gamma functions to express $\Gamma(a+\frac{1}{2})$ and the recursion $\Gamma(1+z) = z\Gamma(z)$ to express $\Gamma(2-2a) = \Gamma(1+(1-2a))$; in the fifth and sixth equations we have used Euler's reflection formula for $\Gamma(1-z)\Gamma(z)$.

[1] Analogously, $I^{+} > 0.5$ (i.e. $\mu_W > 0$) leads to positive skewness $\beta > 0$, while $I^{+} < 0.5$ (i.e. $\mu_W < 0$) leads to $\beta < 0$.

- A joint combination of the scale and location parameters is needed to achieve the limiting cases $\beta = -1$ ($\mu_W < 0, \sigma_W = 0$) and $\beta = 1$ ($\mu_W > 0, \sigma_W = 0$). In particular, from (2.34a),

$$\begin{aligned}\sigma^\alpha = \frac{\mu_W^\alpha}{C_\alpha} &= \mu_W^\alpha \frac{\Gamma(2-\alpha)\cos(\pi\alpha/2)}{1-\alpha} \\ &= \mu_W^\alpha \frac{\Gamma(1+(1-\alpha))}{1-\alpha}\cos(\pi\alpha/2) \\ &= \mu_W^\alpha \Gamma(1-\alpha)\cos(\pi\alpha/2), \end{aligned} \tag{A.3}$$

  where the third equality follows from the recursion $\Gamma(1+z) = z\Gamma(z)$.

A.2 Connection between the SMiN and MaSMiN Models

There is a connection between the SMiN and the MaSMmiN models: the former is a special case of the latter, when the distribution of X is symmetric stable. In fact, using the MaSMiN representation (2.38a), and $\mu_W = 0$ and $\sigma_W > 0$ to yield a symmetric distribution for X, according to Appendix A.1, we have

$$X|(M,S) \sim \mathcal{N}\left(0, \sigma_W^2 S^2\right). \tag{A.4}$$

This means that only the scale variable in the MaSMiN is random, when the distribution of X is symmetric stable. Furthermore, from (2.37), $S^2 = \sum_{j=1}^\infty \Gamma_j^{-2/\alpha} = \sum_{j=1}^\infty \Gamma_j^{-1/\tilde{\alpha}}$, with $\tilde{\alpha} = \alpha/2$, which is itself a form of the PSR (2.29) with $\tilde{W}_j = 1$, or, equivalently, $\tilde{W}_j \sim \mathcal{N}(\tilde{\mu}_W = 1, \tilde{\sigma}_W = 0)$. Then, using (A.3)

$$\begin{aligned} S^2 &\sim \mathcal{S}_{\tilde{\alpha}}\left((\Gamma(1-\tilde{\alpha}))^{1/\tilde{\alpha}}(\cos(\pi\tilde{\alpha}/2))^{1/\tilde{\alpha}}, 1, 0\right) \\ &\overset{\mathcal{D}}{=} (\Gamma(1-\alpha/2))^{2/\alpha}\tilde{S}^2, \end{aligned}$$

where $\tilde{S}^2 \sim \mathcal{S}_{\alpha/2}\left((\cos(\pi\alpha/4))^{2/\alpha}, 1, 0\right)$, and the equality in distribution is simply based on scaling. Combining this consideration with (A.2), we can rewrite (A.4) as

$$\begin{aligned} X|\tilde{S} &\sim \mathcal{N}(0, \sigma_W^2(\Gamma(1-\alpha/2))^{2/\alpha}\tilde{S}^2) \overset{\mathcal{D}}{=} \mathcal{N}(0, 2\sigma^2\tilde{S}^2), \\ \tilde{S} > 0 &\sim \mathcal{S}_{\alpha/2}\left((\cos(\pi\alpha/4))^{2/\alpha}, 1, 0\right), \end{aligned}$$

which is precisely the SMiN representation of X. This observation reflects in the fact that there is a parallel in the inference schemes for these two models, as explained in Sections 3.3.2 and 3.3.4.

A.3 Asymptotic Relation between LVs in the MaSMiN Model

The MaSMiN model more traditionally known in the literature, see [McNeil et al., 2015, p. 77], is of the form

$$\begin{aligned} X|S &\sim \mathcal{N}\left(m(S), \tilde{\sigma}^2 S^2\right), \\ S > 0 &\sim \mathcal{D}_S, \end{aligned} \tag{A.5}$$

where $m : [0, \infty) \to \mathbb{R}$ is a measurable function. For example, if $m(s) = \tilde{\mu}s^2$, $\tilde{\mu} \in \mathbb{R}$, then the mean and variance of the conditionally Gaussian distribution are affine transformation of the same variable and we can write[2]

$$X|S \sim \mathcal{N}\left(\tilde{\mu}S^2, \tilde{\sigma}^2 S^2\right).$$

Observe that this is different from our framework: the variables M and S in (2.36)-(2.37) are dependent, because they involve series of transformations of the same RVs $\{\Gamma_j\}$, but it is not possible to give an explicit function that maps M to S (or vice versa). However, numerical results by Lemke [2014] and Lemke et al. [2015] show that, when the stable random variable $X \sim \mathcal{S}_\alpha(\sigma, \beta, 0)$ takes extremely large values, it is possible identify the approximate relationship

$$S^2 \approx M^2,$$

or, equivalently $M \approx S$. This implies that, in the asymptotic regime $|X| \gg 1$, model (2.38a) becomes

$$X|S \overset{\text{approx}}{\sim} \mathcal{N}(\mu_W S, \sigma_W^2 S^2),$$

which is a special case of (A.5), with $m(s) = \mu_W s$ and $\tilde{\sigma} = \sigma_W$. Hence, in inference procedures, large realizations of X might be treated nearly exactly with a simplified model, an idea that has been developed in Lemke [2014] and Lemke et al. [2015]. However, in the following, we do not consider further such approximation.

[2] One of the most know cases with such representation is when the mixing variable S has a Generalized Inverse Gaussian distribution, in which case X belongs to the class of Generalized Hyperbolic distributions.

Appendix B

Properties for the Proofs of the CLT for $R_{(c,\infty)}$ and the CDF Bounds

B.1 Properties of Characteristic Functions

Recall that the CF of a RV X is defined as $\phi_X(s) := \mathbb{E}[\exp(isX)]$.

We first remark that, for a fixed argument, any CF can be expressed in polar form as $z = re^{i\theta}$, with $r > 0$ and $\theta \in \mathbb{R}$. This polar form is not unique, because integer multiples of 2π can be added to θ, without changing the value of z. However, the polar form is unique if, for example, we require that θ lies in the interval $(-\pi, \pi]$. In this case, the exponential form of z can be inverted, obtaining the *principal value* complex logarithm of the CF, $\log(z) = \log(r) + i\theta$. Although not always necessary, when in this thesis we refer to logarithms of CFs, these are meant to be the principal value complex logarithms.

Moreover, the following Lemmas, listing some properties of CFs (see e.g. [Feller, 1966, pp. 499-500]), prove useful.

Lemma 9. Let $\phi_X(s)$ be the CF of a RV X with distribution F, a, $b \in \mathbb{R}$. Then ϕ is continuous and

(i) $\phi(0) = 1$ and $|\phi(s)| \leq 1$, for all $s \in \mathbb{R}$;

(ii) the RV $aX + b$ has CF $\mathbb{E}[\exp(is(aX + b))] = \phi(aX)\exp(isb)$;

(iii) the real part $\mathrm{Re}(\phi)$ is even while the imaginary part $\mathrm{Im}(\phi)$ is odd. The CF ϕ is real iff F is symmetric.

Lemma 10. Let X_1 and X_2 be two independent RVs. Then their sum $X = X_1 + X_2$ has CF $\phi_X(s) = \phi_{X_1}(s)\phi_{X_2}(s)$.

B.2 Radius of Convergence of the Taylor Series of $\phi_{Y_1}(s)$

In order to prove that $\phi_{Y_1}(s)$ admits Taylor series expansion around $s = 0$ in Lemma 3, it is sufficient to prove that it has non-zero radius of convergence. For this, it is sufficient to show that (5.46) holds, a condition that we report here for easy reference:

$$\limsup_{k\to\infty} \frac{1}{k}|\mathbb{E}[Y_1^k]|^{1/k} < \infty,$$

where $\mathbb{E}[Y_1^k]$ is computed as in (5.16) and we recall that we consider here $W_1 \sim \mathcal{N}(\mu_W, \sigma_W^2)$.

We report a number of lemmas useful to bound $|\mathbb{E}[W_1^k]|$ that is part of $|\mathbb{E}[Y_1^k]|$ above. This is part of work, even if we are aware that a number of alternative proofs related to analyticity of the Gaussian CF might be available in the literature.

Lemma 11. Let us define $g_k := \mathbb{E}[W_1^k]/\mathbb{E}[W_1^2]^{k/2}$, $k \in \mathbb{N}$, with $W_1 \sim \mathcal{N}(\mu_W, \sigma_W^2)$. Then

$$g_k = g_k(\lambda) = \frac{\sum_{j=0}^{k} \binom{k}{j} \mathcal{E}(k-j)(k-j-1)!!\lambda^j}{\left(\sum_{j=0}^{k} \binom{k}{j} \lambda^{2j}\right)^{1/2}}, \tag{B.1}$$

where $n!!$ denotes the double factorial of $n \in \mathbb{N}$, $\lambda := \mu_W/\sigma_W$ and the indicator function $\mathcal{E}(n) := \mathbb{1}(\text{mod}(n, 2) = 0)$ is 1 if n is even, 0 otherwise.

Proof: Defining $\Omega := W_1 - \mu_W$ to be the centred version of W_1, such that $\mathbb{E}\left[\Omega^k\right] = \sigma_W^k\,(k-1)!!\mathcal{E}(k)$, see Winkelbauer [2012], the coefficients g_k become

$$g_k = \frac{\mathbb{E}[(\Omega + \mu_W)^k]}{\left(\mu_W^2 + \sigma_W^2\right)^{k/2}} = \frac{\mathbb{E}[(\Omega + \mu_W)^k]}{\sigma_W^k\,(\lambda^2+1)^{k/2}},$$

where $\lambda = \mu_W/\sigma_W$. If we take the binomial expansion of the numerator and the denominator of the above

$$g_k = \frac{\sum_{j=0}^{k} \binom{k}{j} \mathbb{E}\left[\Omega^{k-j}\right] \mu_W^j}{\sigma_W^k \left(\sum_{j=0}^{k} \binom{k}{j} \lambda^{2j}\right)^{1/2}},$$

from which the statement follows by substituting the expression of the $(k-j)$-th moment of Ω, simplifying σ_W^k and recognizing λ^j in the numerator. ■

To show that the coefficients g_k are bounded $\forall k \geq 3$, we first prove the following preliminary result.

Lemma 12. For any even $k \geq 3$ and even $j \leq k$

$$\binom{k}{j}\frac{(k-j-1)!!}{(k-1)!!} \leq \sqrt{\binom{k}{j}}, \tag{B.2}$$

and, for any odd $k \geq 3$ and $j \leq k$

$$\binom{k}{j}\frac{(k-j-1)!!}{k!!} \leq \sqrt{\binom{k}{j}}. \tag{B.3}$$

Proof: Let us start with even k and j and simplify the left hand side of (B.2)

$$\binom{k}{j}\frac{(k-j-1)!!}{(k-1)!!} = \frac{k!}{(k-j)!j!}\frac{(k-j-1)!!}{(k-1)!!} = \frac{k!!}{(k-j)!!j!}.$$

We then show that the square of the left hand side of (B.2), divided by the square of its right hand side is less than unity

$$\begin{aligned}\frac{k!!^2(k-j)!j!}{(k-j)!!^2j!^2k!} &= \frac{k!!^2(k-j)!}{(k-j)!!^2j!k!} \\ &= \frac{k!!(k-j-1)!!}{(k-j)!!j!(k-1)!!} \\ &= \frac{k(k-2)\cdots(k-j+2)}{(k-1)(k-3)\cdots(k-j+1)j!} \\ &\leq \frac{2^{j/2}}{j!} \leq 1,\end{aligned}$$

where in the second equality we use the fact that $l!/l!! = (l-1)!!$; in the third equality we rearrange the terms and we elide the common ones; finally, in the first inequality we use the fact that the fraction $l/(l-1) \leq 2$, $\forall l \geq 2$, and both the numerator and the denominator have $j/2$ terms, and the last inequality is well known.

When k and j are odd

$$\binom{k}{j}\frac{(k-j-1)!!}{k!!} = \frac{k!}{(k-j)!j!}\frac{(k-j-1)!!}{k!!} = \frac{(k-1)!!}{(k-j)!!j!}$$

and, in (B.3), the square of the left hand side divided by the square of the right hand side is less than unity

$$\begin{aligned}\frac{(k-1)!!^2(k-j)!j!}{(k-j)!!^2j!^2k!} &= \frac{(k-1)!!^2(k-j)!}{(k-j)!!^2j!k!} \\ &= \frac{(k-1)!!(k-j-1)!!}{(k-j)!!j!k!!} \\ &= \frac{(k-1)(k-3)\cdots(k-j+2)}{k(k-2)\cdots(k-j+1)j!} \leq 1\end{aligned}$$

since for all terms in the numerator there is a larger term in the denominator. ■

Combining Lemma 11 and Lemma 12, we have the following.

Lemma 13. Let g_k, $k \geq 3$, be defined as in (B.1). Then $0 \leq g_k \leq (k-1)!!$ for even k, and $|g_k| \leq k!!$ for odd k, and $\mathrm{sgn}(g_k) = \mathrm{sgn}(\lambda)$, where $\lambda = \mu_W/\sigma_W$.

Proof: For even k

$$\frac{1}{(k-1)!!}g(k) \leq \frac{\sum_{j=0}^{k}\sqrt{\binom{k}{j}}\mathcal{E}(k-j)\lambda^j}{\left(\sum_{j=0}^{k}\binom{k}{j}\lambda^{2j}\right)^{1/2}} \leq 1 \tag{B.4}$$

where the first inequality follows from (B.2), and the second inequality follows from the unity $\frac{|u|}{\sqrt{\|u\|_2^2+\|v\|_2^2}} \leq 1$, where we use $u = \sum_{j=0}^{k}\sqrt{\binom{k}{j}}\mathcal{E}(k-j)\lambda^j$, $v = \sum_{j=0}^{k}\sqrt{\binom{k}{j}}\mathcal{O}(k-j)\lambda^j$, and the indicator function $\mathcal{O}(n) := \mathbb{1}(\mathrm{mod}(n,2)=1) = 1-\mathcal{E}(n)$ is 1 if (n) is odd, 0 otherwise.

Similarly, for odd k

$$\frac{1}{k!!}|g(k)| \leq \frac{\left|\sum_{j=0}^{k}\sqrt{\binom{k}{j}}\mathcal{E}(k-j)\lambda^j\right|}{\left(\sum_{j=0}^{k}\binom{k}{j}\lambda^{2j}\right)^{1/2}} \leq 1.$$

The sign of the coefficients g_k is clear from (B.1). ■

The ratio between $g_k(\lambda)$ (B.1) and either $(k-1)!!$, for even k, or $k!!$, for odd k, is shown in Figure B.1. Thus, from Lemma 13, we have that, when $W \sim \mathcal{N}(\mu_W, \sigma_W^2)$,

$$|\mathbb{E}[W_1^k]| \leq \begin{cases}(\mu_W^2+\sigma_W^2)^{k/2}(k-1)!!, & \text{if } k \text{ is even,}\\ (\mu_W^2+\sigma_W^2)^{k/2}k!!, & \text{if } k \text{ is odd.}\end{cases}$$

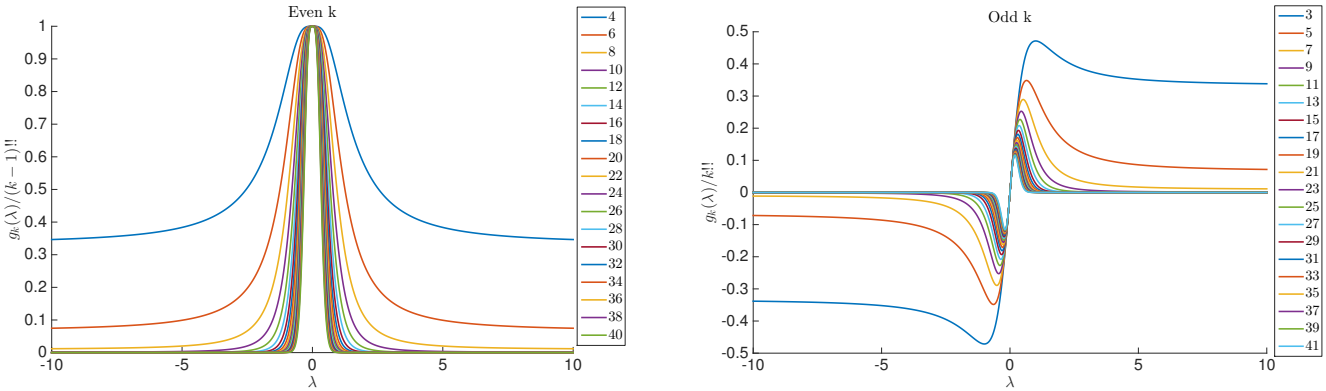


Figure B.1 Left: $g_k(\lambda)/(k-1)!!$, for even $k \geq 3$; right: $g_k(\lambda)/k!!$, for odd $k \geq 3$.

Given the relation $k! = k!!(k-1)!!$, the following asymptotic holds for the double factorial

$$k!! \sim (k-1)!! \sim \sqrt{k!}, \qquad k \to \infty.$$

We can compute R, the desired radius of convergence of the Taylor series of $\phi_{Y_1}(s)$ as

$$\begin{aligned} R^{-1}/e &= \limsup_{k\to\infty} \frac{1}{k} |\mathbb{E}[W_1^k]|^{1/k} \frac{1}{(d-c)^{1/k}} \left(\frac{\alpha}{k-\alpha}\right)^{1/k} \left(c^{\frac{\alpha-k}{\alpha}} - d^{\frac{\alpha-k}{\alpha}}\right)^{1/k} \\ &\leq \limsup_{k\to\infty} \frac{1}{k} (\mu_W^2 + \sigma_W^2)^{1/2} (k!)^{1/2k} \left(\frac{\alpha}{k-\alpha}\right)^{1/k} \left(c^{\frac{\alpha-k}{\alpha}}\right)^{1/k} \\ &= \limsup_{k\to\infty} \frac{1}{k} (\mu_W^2 + \sigma_W^2)^{1/2} \left(\frac{k}{e}\right)^{1/2} c^{-1/\alpha} \\ &= 0, \end{aligned} \tag{B.5}$$

where, in the inequality the supremum of $c^{(\alpha-k)/\alpha} - d^{(\alpha-k)/\alpha}$ is obtained when $d \to \infty$, while in the second equality we have solved the indeterminate limit $[0]^0$ by first exponentiating and then using L'Hôpital's rule. Thus (B.5) implies that the Taylor series expansion of $\phi_{Y_1}(s)$ $s \in \mathbb{R}$ converges absolutely for $s \in \mathbb{R}$.

B.3 Uniform Convergence of Series of Functions

Here we report the definitions of uniform convergence of a series of functions

$$S(x) := \sum_{k=1}^{\infty} a_k(x), \qquad x \in \mathcal{X} \tag{B.6}$$

and some related properties, referring to [Rudin, 1976, Chpater 7]. For our purposes, in the proof of Lemma 3, it is sufficient to consider complex valued functions $(a_k(x) \in \mathbb{C})$ of a real variable $x \in \mathcal{X} \subseteq \mathbb{R}$.

Definition B.3.1. The series of functions (B.6) is uniformly convergent if for every $\epsilon > 0$ there is an N_ϵ independent of x, such that, for all $N > N_\epsilon$ and $x \in \mathcal{X}$

$$|S(x) - S_N(x)| < \epsilon,$$

where

$$S_N(x) := \sum_{k=1}^{N} a_k(x), \qquad x \in \mathcal{X} \tag{B.7}$$

The above definition is equivalent to the fact that the sequence of partial sums $S_N(x)$ converges uniformly on $\mathcal{X}$. The following sufficient but not necessary condition is useful to establish if a series of functions is uniformly converging.

Theorem B.3.1. Suppose that $\{a_k\}$ is a sequence of functions defined on $\mathcal{X}$, and suppose

$$|a_k(x)| \leq M_k, \qquad x \in \mathcal{X}, k = 1, 2, \ldots$$

Then $\sum_{k=1}^{\infty} a_k(x)$ converges uniformly on $\mathcal{X}$ if $\sum_{k=1}^{\infty} M_k$ converges.

Finally, we are interested in the following property.

Lemma 14. If $\{a_k\}$ is a sequence of continuous function on $\mathcal{X}$, and the series $S(x)$ in (B.6) is uniformly converging on $\mathcal{X}$, then $S(x)$ is continuous on $\mathcal{X}$, that is

$$\lim_{x \to x_0} S(x) = \lim_{x \to x_0} \sum_{k=1}^{\infty} a_k(x) = \sum_{k=1}^{\infty} \lim_{x \to x_0} a_k(x) = S(x_0). \tag{B.8}$$

The above definitions and properties are used in equation (5.50) in the proof of Lemma 3. In particular, (5.50) is obtained as the limit, for a fixed value of $s \in \mathbb{R}$, and as $d \to \infty$, of the series of functions in (5.48)

$$S(d, s) = \sum_{k=3}^{\infty} r_k s^k, \tag{B.9}$$

where the variable $d > \max\{1, c\}$ is implicit in the coefficients r_k, $k \geq 3$, defined in (5.49) and reported here for easy reference:

$$r_k = \frac{i^k}{k!}\mathbb{E}[W_1^k]\frac{\alpha}{\alpha - k}\left(d^{(\alpha-k)/\alpha} - c^{(\alpha-k)/\alpha}\right)$$

Based on Lemma 14, we can take the term-by-term limit of (B.9), because the coefficients r_k are a continuous function of d, and $S(d, s)$ is uniformly convergent in d thanks to Theorem B.3.1. In fact, for a fixed s, its summands $r_k s^k$ are uniformly bounded in absolute value because the coefficients r_k are uniformly bounded,

$$\sup_{d>\max\{1,c\}} |r_k| = \frac{1}{k!}|\mathbb{E}[W_1^k]|\frac{\alpha}{k - \alpha}c^{(\alpha-k)/\alpha} := l_k.$$

Moreover, $\sum_{k\geq 3}^{\infty} l_k s^k$ is a power series whose radius of convergence can be proved to be positive, using similar arguments to those in Appendix B.2.

B.4 Properties of Gamma Functions

Two complex-analytic functions appear repeatedly in our analysis are the lower incomplete gamma function,

$$\gamma(s, x) := \int_0^x t^{s-1}e^{-t}\,\mathrm{d}t, \qquad -s \notin \mathbb{N}, x > 0, \tag{B.10}$$

and the upper incomplete gamma function,

$$\Gamma(s, x) := \int_x^{\infty} t^{s-1}e^{-t}\,\mathrm{d}t, \qquad -s \notin \mathbb{N}, x > 0. \tag{B.11}$$

Then, for any $x > 0$,

$$\Gamma(s) = \gamma(s, x) + \Gamma(s, x), \qquad -s \notin \mathbb{N}, \tag{B.12}$$

is the regular (complete) gamma function.

B.4.1 Inequalities for $\gamma(s, x)$

We list some simple properties of the lower incomplete gamma function (B.10) that have been used throughout the thesis.

Lemma 15. *[Neuman, 2013, Theorem 4.1]* For all $x > 0$ and $s \in (0, 1]$

$$\gamma(s, x) \leq \frac{x^s}{s(s + 1)}(1 + se^{-x}).$$

Lemma 16. *[Olver et al., 2018, Ineq. (8.10.2)]* For all $x > 0$ and $s > 0$

$$\gamma(s,x) \geq \frac{x^{s-1}}{s}(1-e^{-x}). \tag{B.13}$$

Combining Lemmas 15 and 16 we obtain that

$$\lim_{x\downarrow 0} \frac{\gamma(s,x)}{x^s} = \frac{1}{s}, \tag{B.14}$$

and also the following:

Lemma 17. For all $x > 0$ and $s \in (0,1]$:

$$-\frac{x}{1+s} \leq e^{-x} - \frac{s\gamma(s,x)}{x^s} \leq -\frac{x(1-x)}{2}.$$

Proof: The bounds in Lemmas 15 and 16 immediately give,

$$\frac{1-e^{-x}}{x} \leq \frac{s\gamma(s,x)}{x^s} \leq \frac{1+se^{-x}}{1+s},$$

and subtracting all three sides from e^{-x} and simplifying,

$$\frac{e^{-x}-1}{1+s} \leq e^{-x} - \frac{s\gamma(s,x)}{x^s} \leq \frac{(x+1)e^{-x}-1}{x}. \tag{B.15}$$

Applying the elementary inequality $e^{-x} \geq 1-x$, $x \geq 0$, to the left-hand side gives the lower bound in the statement, and similarly applying the inequality $e^{-x} \leq 1-x+\frac{x^2}{2}$, $x \geq 0$, to the right-hand side gives the corresponding upper bound. ■

B.4.2 Inequalities for $\Gamma(s,x)$

We also make use of the following properties of the upper incomplete gamma function (B.11).

Lemma 18. *[Olver et al., 2018, Eq. (8.6.7)]* For all $s > 0, x > 0$, $\Gamma(s,x)$ admits the representation:

$$\Gamma(s,x) = x^s \int_0^\infty \exp(st - xe^t)\,\mathrm{d}t.$$

Lemma 19. For all $x > s > 0$:

$$\frac{\Gamma(s,x)}{x^s e^{-x}} \leq \frac{1}{x-s}.$$

Proof: From Lemma 18,

$$\frac{\Gamma(s,x)}{x^s e^{-x}} = \int_0^\infty \exp(st - xe^t + x)\,\mathrm{d}t, \tag{B.16}$$

and since $e^t \geq t+1$ for $t \geq 0$,

$$\frac{\Gamma(s,x)}{x^s e^{-x}} \leq \int_0^\infty \exp(-t(x-s))\,\mathrm{d}t = \frac{1}{x-s}.$$

■

The bound in the last lemma applies when $x > s$. When x may be smaller than s we have a somewhat weaker bound, but this time uniformly in x:

Lemma 20. Let $\delta \in (0,1)$ be arbitrary. For all $s > 0$ and all $x \geq \delta s$:

$$\frac{\Gamma(s,x)}{x^s e^{-x}} \leq \sqrt{\frac{2\pi}{\delta s}} \exp\left[\frac{\delta s}{2}(\delta^{-1}-1)^2\right].$$

Proof: Starting again with (B.16) and noting that the integrand is decreasing in x, we have that, for $x \geq \delta s$,

$$\frac{\Gamma(s,x)}{x^s e^{-x}} \leq \int_0^\infty \exp\left[s(t - \delta e^t + \delta)\right]\mathrm{d}t,$$

and since $e^t \geq 1 + t + \frac{t^2}{2}$ for $t \geq 0$,

$$\frac{\Gamma(s,x)}{x^s e^{-x}} \leq \int_0^\infty \exp\left\{-\frac{\delta s}{2}\left[t^2 - 2t(\delta^{-1}-1)\right]\right\}\mathrm{d}t.$$

Completing the square in the exponent,

$$\begin{aligned}\frac{\Gamma(s,x)}{x^s e^{-x}} &\leq \exp\left[\frac{\delta s}{2}(\delta^{-1}-1)^2\right]\int_0^\infty \exp\left\{-\frac{\delta s}{2}\left[t-(\delta^{-1}-1)\right]^2\right\}\mathrm{d}t \\ &\leq \sqrt{\frac{2\pi}{\delta s}}\exp\left[\frac{\delta s}{2}\left(\frac{1}{\delta}-1\right)^2\right] F_Z\left((\delta^{-1}-1)(\delta s)^{1/2}\right),\end{aligned}$$

where, as in the rest of the paper, $F_Z(\cdot)$ denotes the standard normal CDF, which is, of course, no greater than 1. ■

B.4.3 Asymptotics

When $x \in \mathbb{R}$, the upper incomplete gamma function (B.11) has the following asymptotic behaviour for $x \to \infty$

$$\Gamma(s, x) \sim x^{s-1} e^{-x}. \tag{B.17}$$

This can be proved using series expansions of $\Gamma(s, x)$, see [Abramowitz and Stegun, 1964, eq. (6.5.32), p. 263] or Olver et al. [2018].

B.4.4 Derivatives

Finally, by the fundamental theorem of calculus and the definitions (B.10) and (B.11), we have

$$\frac{\mathrm{d}\gamma(s, x)}{\mathrm{d}x} = -\frac{\mathrm{d}\Gamma(s, x)}{\mathrm{d}x} = x^{s-1} e^{-x}. \tag{B.18}$$

... In reality, the number of drawings is infinite. No decision is final; all branch into others. The inexperienced assume that infinite drawings require infinite time; actually, all that is required is that time is infinitely subdivisible, as in the famous parable of the Race with the Tortoise ...

The lottery of Babylon, *Labyrinths*, Jorge Luis Borges

Bibliography

Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*. Applied mathematics series. Dover Publications. 202

Achim, A., Bezerianos, A., and Tsakalides, P. (2001). Novel Bayesian multiscale method for speckle removal in medical ultrasound images. *IEEE Transactions on Medical Imaging*, 20(8):772–783. 3

Achim, A., Kuruoğlu, E. E., and Zerubia, J. (2006). SAR image filtering based on the heavy-tailed Rayleigh model. *IEEE Transactions on Image Processing*, 15(9):2686–2693. 3

Ament, S. and O'Neil, M. (2018). Accurate and efficient numerical calculation of stable densities via optimized quadrature and asymptotics. *Statistics and Computing*, 28(1):171–185. 6

Amini, A. and Unser, M. (2014). Sparsity and infinite divisibility. *IEEE Transactions on Information Theory*, 60(4):2346–2358. 3, 26

Andrews, D. F. and Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 99–102. 32

Andrieu, C. and Roberts, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *The Annals of Statistics*, 37(2):697–725. 8, 57

Andrieu, C. and Thoms, J. (2008). A tutorial on adaptive MCMC. *Statistics and computing*, 18(4):343–373. 50

Applebaum, D. (2004). *Lévy Processes and Stochastic Calculus*. Cambridge studies in advanced mathematics. Cambridge University Press. 23

Asmussen, S. and Rosiński, J. (2001). Approximations of small jumps of Lévy processes with a view towards simulation. *Journal of Applied Probability*, 38(2):482–493. 25

Aström, K. J. and Murray, R. M. (2010). *Feedback systems: an introduction for scientists and engineers*. Princeton University Press. 177

Azzaoui, N. and Clavier, L. (2010). Statistical channel model based on α-stable random processes and application to the 60 GHz ultra wide band channel. *IEEE Transactions on Communications*, 58(5):1457–1467. 3, 39

Barndorff-Nielsen, O., Mikosch, T., and Resnick, S. (2001). *Lévy Processes: Theory and Applications*. Birkhäuser Boston. 23, 26

Barndorff-Nielsen, O. E. and Shephard, N. (2001). Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):167–241. 26

Beaumont, M. A. (2003). Estimation of population growth or decline in genetically monitored populations. *Genetics*, 164(3):1139–1160. 57

Bentkus, V., Juozulynas, A., and Paulauskas, V. (2001). Lévy–LePage series representation of stable vectors: convergence in variation. *Journal of Theoretical Probability*, 14(4):949–978. 165

Berger, J. O. (2013). *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media. 44

Bertoin, J. (1998). *Lévy Processes*. Cambridge Tracts in Mathematics. Cambridge University Press. 23

Betancourt, M. (2017). A conceptual introduction to Hamiltonian Monte Carlo. *arXiv preprint arXiv:1701.02434*. 50

Billingsley, P. (2008). *Probability and measure*. John Wiley & Sons. 23

Borodin, A. N. and Salminen, P. (2012). *Handbook of Brownian motion-facts and formulae*. Birkhäuser. 23

Brockwell, P. J. (2001). Lévy-driven CARMA processes. *Annals of the Institute of Statistical Mathematics*, 53(1):113–124. 26, 177

Brockwell, P. J. (2004). Representations of continuous-time ARMA processes. *Journal of Applied Probability*, 41A:375–382. 26, 177

Brockwell, P. J. and Lindner, A. (2009). Existence and uniqueness of stationary Lévy-driven CARMA processes. *Stochastic Processes and their Applications*, 119(8):2660–2681. 26, 177

Buckle, D. (1994). The study of a function relating to stable distributions. *Statistics & Probability Letters*, 20(2):85–90. 30, 74

Buckle, D. J. (1995). Bayesian inference for Stable distributions. *Journal of the American Statistical Association*, 90(430):pp. 605–613. 8, 30, 61, 62, 72, 73, 74, 76, 77, 78, 81, 82, 84, 85, 95, 96, 100, 183

Byczkowski, T., Nolan, J. P., and Rajput, B. (1993). Approximation of multidimensional stable densities. *Journal of Multivariate Analysis*, 46(1):13–31. 20

Cappé, O., Godsill, S. J., and Moulines, E. (2007). An overview of existing methods and recent advances in sequential Monte Carlo. *Proceedings of the IEEE*, 95(5):899–924. 9

Carrillo, R. E., Ramirez, A. B., Arce, G. R., Barner, K. E., and Sadler, B. M. (2016). Robust compressive sensing of sparse signals: a review. *EURASIP Journal on Advances in Signal Processing*, 2016(1):108. 3

Chambers, J. M., Mallows, C. L., and Stuck, B. W. (1976). A Method for Simulating Stable Random Variables. *Journal of the American Statistical Association*, 71(354):340–344. 8, 16

Chen, X., Wang, Z. J., and McKeown, M. J. (2010). Asymptotic analysis of robust lassos in the presence of noise with large variance. *IEEE Transactions on Information Theory*, 56(10):5131–5149. 3

Christensen, H. L., Murphy, J., and Godsill, S. J. (2012). Forecasting high-frequency futures returns using online Langevin dynamics. *IEEE Journal of Selected Topics in Signal Processing*, 6(4):366–380. 177

Clauset, A., Shalizi, C. R., and Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM review*, 51(4):661–703. 3

de Freitas, M. L., Egan, M., Clavier, L., Goupil, A., Peters, G. W., and Azzaoui, N. (2017). Capacity bounds for additive symmetric α-stable noise channels. *IEEE Transactions on Information Theory*, 63:5115–5123. 3

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38. 60

Devroye, L. (1986). *Non-uniform random variate generation*. Springer. 45, 86, 94

Doucet, A., Godsill, S. J., and Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10(3):197–208. 9, 10

Doucet, A. and Johansen, A. M. (2011). A tutorial on particle filtering and smoothing: fifteen years later. 9

Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. (1987). Hybrid Monte Carlo. *Physics letters B*, 195(2):216–222. 50

Dutka, J. (1991). The early history of the factorial function. *Archive for History of Exact Sciences*, 43(3):225–249. 122

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events*. Stochastic Modelling and Applied Probability. Springer-Verlag Berlin Heidelberg. 4, 5, 23, 24

Fama, E. F. (1965). The behavior of stock-market prices. *The Journal of Business*, 38(1):pp. 34–105. 3

Feldheim, E. (1937). *Étude de la stabilité des lois de probabilité*. 19

Feller, W. (1966). *An introduction to probability theory and its applications*. Number v. 2 in Wiley mathematical statistics series. Wiley. 2, 4, 5, 6, 8, 14, 32, 84, 109, 110, 111, 118, 119, 122, 193

Franklin, G. F., Powell, J. D., Emami-Naeini, A., and Powell, J. D. (1994). *Feedback control of dynamic systems*, volume 3. Addison-Wesley Reading, MA. 177

Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410):398–409. 50

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2014). *Bayesian Data Analysis, Third Edition (Chapman & Hall/CRC Texts in Statistical Science)*. Chapman and Hall/CRC, London, third edition. 64

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-6(6):721–741. 50

Gilks, W. R., Best, N., and Tan, K. (1995a). Adaptive rejection Metropolis sampling within Gibbs sampling. *Applied Statistics*, pages 455–472. 50

Gilks, W. R., Richardson, S., and Spiegelhalter, D. (1995b). *Markov chain Monte Carlo in practice*. CRC press. 48

Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 41(2):pp. 337–348. 45, 96

Gill, P. E., Murray, W., and Wright, M. H. (1981). *Practical optimization*. Academic Press, New York. 77

Gnedenko, B. V. and Kolmogorov, A. N. (1968). *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley series in statistics. Addison-Wesley. 4, 5, 6, 15

Godsill, S. (1999). MCMC and EM-based methods for inference in heavy-tailed processes with alpha-stable innovations. In *Higher-Order Statistics, 1999. Proceedings of the IEEE Signal Processing Workshop on*, pages 228–232. 9, 63, 64, 65

Godsill, S. (2000). Inference in symmetric alpha-stable noise using MCMC and the slice sampler. In *Acoustics, Speech, and Signal Processing, 2000. ICASSP '00. Proceedings. 2000 IEEE International Conference on*, volume 6, pages 3806–3809 vol.6. 9, 65

Godsill, S. J. and Kuruoğlu, E. E. (1999). Bayesian inference for time series with heavy-tailed symmetric α-stable noise processes. In *Applications of Heavy Tailed Distributions in Economics, Engineering and Statistics*, Washington DC, USA. 9, 63, 64, 65

Grabchak, M. and Samorodnitsky, G. (2010). Do financial returns have finite or infinite variance? a paradox and an explanation. *Quantitative Finance*, 10(8):883–893. 15

Haario, H., Saksman, E., and Tamminen, J. (2001). An adaptive metropolis algorithm. *Bernoulli*, 7(2):223–242. 50

Harvey, A. C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press. 26

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):pp. 97–109. 48

Hida, T. (1980). Brownian motion. In *Brownian Motion*, pages 44–113. Springer. 23

Hoffmann, J. and Jørgensen, J. (1994). Stable densities. *Theory of Probability & Its Applications*, 38(2):350–355. 6

Horn, R. A., Horn, R. A., and Johnson, C. R. (1990). *Matrix analysis*. Cambridge university press. 177

Janicki, A. and Kokoszka, P. (1992). On the rate of convergence of LePage type series to finite dimensional distributions of Lévy motion. *preprint*. 39

Janicki, A. and Weron, A. (1994). *Simulation and Chaotic Behavior of α-stable Stochastic Processes*. HSC Books. Hugo Steinhaus Center, Wroclaw University of Technology. 23, 24, 39

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *ASME Journal of Basic Engineering*. 9

Katz, R. W. and Brown, B. G. (1992). Extreme events in a changing climate: Variability is more important than averages. *Climatic Change*, 21(3):289–302. 3

Katz, R. W., Parlange, M. B., and Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources*, 25(8,Äì12):1287 – 1304. 3

Khinchine, A. Y. and Lévy, P. (1936). Sur les lois stables. *CR Acad. Sci. Paris*, 202:374–376. 13

Kinderman, A. J. and Monahan, J. F. (1977). Computer generation of random variables using the ratio of uniform deviates. *ACM Transactions on Mathematical Software (TOMS)*, 3(3):257–260. 66

Kingman, J. (1992). *Poisson Processes*. Oxford Studies in Probability. Clarendon Press. 106

Kloeden, P. E. and Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations*, volume 23. Springer. 27

Koutrouvelis, I. A. (1980). Regression-type estimation of the parameters of Stable laws. *Journal of the American Statistical Association*, 75(372):pp. 918–928. 8

Kuruoğlu, E. E. (2001). Density parameter estimation of skewed α-stable distributions. *Signal Processing, IEEE Transactions on*, 49(10):2192–2201. 8

Ledoux, M. and Paulauskas, V. (1996). A rate of convergence in the Poissonian representation of stable distributions. *Lithuanian Mathematical Journal*, 36(4):388–399. 39, 154, 157

Lehmann, E. L. and Casella, G. (2006). *Theory of point estimation*. Springer Science & Business Media. 44

Lemke, T. (2014). *Poisson Series Approaches to Bayesian Monte Carlo Inference for Skewed α-Stable Distributions and Stochastic Processes*. PhD thesis, Kaiserslautern University of Technology. 9, 10, 39, 67, 68, 70, 105, 110, 113, 159, 162, 165, 166, 169, 192

Lemke, T. and Godsill, S. J. (2011). Enhanced Poisson sum representation for alpha-stable processes. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pages 4100–4103. 10, 39, 162, 165

Lemke, T. and Godsill, S. J. (2012). Linear Gaussian computations for near-exact Bayesian Monte Carlo inference in skewed alpha-stable time series models. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pages 3737–3740. 39, 69, 113

Lemke, T. and Godsill, S. J. (2014). A Poisson series approach to Bayesian Monte Carlo inference for skewed alpha-stable distributions. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 8023–8027. IEEE. 9, 39, 67, 113

Lemke, T. and Godsill, S. J. (2015). Inference for models with asymmetric α-stable noise processes. In Koopman S. J. and Shephard N., editor, *Unobserved Components and Time Series Econometrics*, chapter 9. Oxford University Press, Oxford. 10, 70, 113, 165, 166, 169

Lemke, T., Riabiz, M., and Godsill, S. J. (2015). Fully Bayesian inference for α-stable distributions using a Poisson series representation. *Digital Signal Processing*, 47:96 – 115. 9, 39, 67, 68, 113, 192

Lemons, D. S. and Gythiel, A. (1997). Paul Langevin's 1908 paper"On the Theory of Brownian Motion"["Sur la théorie du mouvement brownien,"CR Acad. Sci.(Paris) 146, 530-533 (1908)]. *American Journal of Physics*, 65:1079–1081. 26, 178

LePage, R. (1981). Multidimensional infinitely divisible variables and processes Part II. In *Probability in Banach Spaces III*, pages 279–284. Springer. 34

LePage, R. (1989). Appendix Multidimensional infinitely divisible variables and processes. Part I: Stable case. In *Probability theory on vector spaces IV*, pages 153–163. Springer. 34

LePage, R., Woodroofe, M., and Zinn, J. (1981). Convergence to a Stable distribution via order statistics. *The Annals of Probability*, 9(4):624–632. 34

Levine, R. A. and Casella, G. (2001). Implementations of the Monte Carlo EM Algorithm. *Journal of Computational and Graphical Statistics*, 10(3):422–439. 60

Lévy, P. (1924). Théorie des erreurs. la loi de gauss et les lois exceptionelles. *Bull. Soc. Math. France*, 52:49–85. 13

Lévy, P. (1925). *Calcul des Probabilités*. PCMI collection. Gauthier-Villars. 13

Liebeherr, J., Burchard, A., and Ciucu, F. (2012). Delay bounds in communication networks with heavy-tailed and self-similar traffic. *IEEE Transactions on Information Theory*, 58(2):1010–1024. 3

Lindeberg, J. W. (1922). Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 15:211–225. 5

Lombardi, M. J. (2007). Bayesian inference for α-stable distributions: A random walk MCMC approach . *Computational Statistics & Data Analysis*, 51(5):2688 – 2700. 8, 70

Lombardi, M. J. and Godsill, S. J. (2004). On-line Bayesian estimation of AR signals in symmetric $\alpha-$stable noise. Econometrics working papers archive, Universita' degli Studi di Firenze, Dipartimento di Statistica, Informatica, Applicazioni "G. Parenti". 10, 70

Lombardi, M. J. and Godsill, S. J. (2006). On-line Bayesian estimation of signals in symmetric α-stable noise. *Signal Processing, IEEE Transactions on*, 54(2):775–779. 3, 10, 70

MacKay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge University Press. 91

Magnus, J. R. and Neudecker, H. (1980). The elimination matrix: some lemmas and applications. *SIAM Journal on Algebraic Discrete Methods*, 1(4):422–449. 176

Mandelbrot, B. (1963). New methods in statistical economics. *Journal of Political Economy*, 71(5):pp. 421–440. 3

Martino, L., Read, J., and Luengo, D. (2012). Improved adaptive rejection Metropolis sampling algorithms. *arXiv preprint arXiv:1205.5494*. 50

Massey, F. J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, 46(253):68–78. 117

Masuda, H. (2004). On multidimensional Ornstein-Uhlenbeck processes driven by a general Lévy process. *Bernoulli*, 10(1):97–120. 26

McCulloch, J. H. (1986). Simple consistent estimators of stable distribution parameters. *Communications in Statistics - Simulation and Computation*, 15(4):1109–1136. 8

McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative risk management: Concepts, techniques and tools*. Princeton university press. 192

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092. 48

Minka, T. (2005). Divergence measures and message passing. Technical report, Technical report, Microsoft Research. 116

Mittnik, S., , Rachev, S. T., Doganoglu, T., and Chenyao, D. (1999). Maximum likelihood estimation of stable Paretian models. *Mathematical and Computer Modelling*, 29(10-12):275–293. 8

Mitzenmacher, M. (2004). A brief history of generative models for power law and lognormal distributions. *Internet mathematics*, 1(2):226–251. 3

Moler, C. and Van Loan, C. (2003). Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM review*, 45(1):3–49. 175

Muller, P. (1991). A generic approach to posterior integration and Gibbs sampling. Technical report, Purdue University, West Lafayette, Indiana. 51

Müller, P. (1992). Alternatives to the Gibbs sampling scheme. 51

Murphy, J. K. (2014). *Hidden states, hidden structures: Bayesian learning in time series models*. PhD thesis, University of Cambridge. 177

Neal, R. M. et al. (2011). MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 2(11). 50

Neuman, E. (2013). Inequalities and bounds for the incomplete gamma function. *Results in Mathematics*, 63(3):1209–1214. 199

Nikias, C. L. and Shao, M. (1995). *Signal processing with alpha-stable distributions and applications*. Adaptive and learning systems for signal processing, communications, and control. Wiley. 3

Nolan, J. (2001). Maximum likelihood estimation of stable parameters. *Lévy processes: Theory and applications*, pages 379–400. 8

Nolan, J. (2018a). [online] http://fs2.american.edu/jpnolan/www/stable/StableBibliography.pdf. 4, 7

Nolan, J. (2018b). STABLE package for Matlab. [online] http://www.robustanalysis.com/. 20, 30

Nolan, J. P. (1997). Numerical calculation of Stable densities and distribution functions. *Comm. Statist. Stochastic Models*, 13(4):759–774. 6, 14, 30

Nolan, J. P. (1998). Parameterizations and modes of Stable distributions . *Statistics & Probability Letters*, 38(2):187 – 195. 14, 16

Nolan, J. P. (2008). An overview of multivariate stable distributions. 19, 20

Nolan, J. P. (2013). Multivariate elliptically contoured stable distributions: theory and estimation. *Computational Statistics*, 28(5):2067–2089. 8

Nolan, J. P., Panorska, A. K., and McCulloch, J. H. (2001). Estimation of stable spectral measures. *Mathematical and Computer Modelling*, 34(9-11):1113–1122. 8

Øksendal, B. (2003). *Stochastic Differential Equations: An Introduction with Applications*. Springer. 26

Olver, F. W. J., Olde Daalhuis, A. B., Lozier, D. W., Schneider, B. I., Boisvert, R. F., Clark, C. W., Miller, B. R., and B. V. Saunders, e. (2018). *NIST Digital Library of Mathematical Functions*. Release 1.0.17 of 2017-12-22. [online] http://dlmf.nist.gov/. 127, 160, 200, 202

ONeill, P. D., Balding, D. J., Becker, N. G., Eerola, M., and Mollison, D. (2000). Analyses of infectious disease data from household outbreaks by Markov chain Monte Carlo methods. *Journal of The Royal Statistical Society Series C-applied Statistics*, 49:517–542. 57

Paulson, A. S., Holcomb, E. W., and Leitch, R. A. (1975). The estimation of the parameters of the stable laws. *Biometrika*, 62(1):163–170. 8

Peters, G. W., Kannan, B., Lasscock, B., Mellen, C., and Godsill, S. (2011). Bayesian cointegrated vector autoregression models incorporating alpha-stable noise for inter-day price movements via approximate bayesian computation. *Bayesian Anal.*, 6(4):755–792. 8, 70

Peters, G. W., Sisson, S. A., and Fan, Y. (2012). Likelihood-free Bayesian inference for α-stable models. *Computational Statistics & Data Analysis*, 56(11):3743–3756. 8, 70

Qiou, Z. and Ravishanker, N. (1998a). Bayesian inference for time series with stable innovations. *Journal of Time Series Analysis*, 19(2):235–249. 8, 62, 63

Qiou, Z. and Ravishanker, N. (1998b). Bayesian inference for vector ARMA models with stable innovations. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 459–475. 65, 66, 67

Rachev, S. and Mittnik, S. (2000). Stable Paretian models in finance. 3

Rachev, S. T., Klebanov, L. B., Stoyanov, S. V., and Fabozzi, F. J. (2013). *Probability Distances and Probability Metrics: Definitions*, pages 11–31. Springer New York, New York, NY. 115

Ravishanker, N. and Qiou, Z. (1998). Bayesian inference for time series with infinite variance stable innovations. *R. Adler, R. Feldman, and M. Taqqu, editors, A Practical Guide to Heavy Tails*, pages 259–282. 65, 66

Ravishanker, N. and Qiou, Z. (1999). Monte Carlo EM estimation for multivariate stable distributions. *Statistics & probability letters*, 45(4):335–340. 65, 66

Revuz, D. and Yor, M. (2013). *Continuous martingales and Brownian motion*, volume 293. Springer Science & Business Media. 23

Riabiz, M., Ardeshiri, T., and Godsill, S. (2017a). A central limit theorem with application to inference in α-stable regression models. In *Proceedings of the Time Series Workshop at NIPS 2016*, volume 55 of *Proceedings of Machine Learning Research*, pages 70–82, Barcelona, Spain. PMLR. 105, 110, 121, 125, 127

Riabiz, M., Ardeshiri, T., and Godsill, S. (2017b). Convergence results for tractable inference in α-stable stochastic processes. In *2017 22nd International Conference on Digital Signal Processing (DSP)*, pages 1–5. 105, 109, 116

Riabiz, M., Ardeshiri, T., Kontoyiannis, I., and Godsill, S. (2018a). Nonasymptotic Gaussian Approximation for Linear Systems with Stable Noise. *arXiv preprint arXiv:1802.10065v3*. 105, 110, 118, 121, 125, 131, 144, 148

Riabiz, M., Ardeshiri, T., Kontoyiannis, I., and Godsill, S. J. (2017c). Simulated convergence rates with application to an intractable α-stable inference problem. In *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP) (IEEE CAMSAP 2017)*, Curaccao, Netherlands Antilles. 118, 131, 141, 144, 153

Riabiz, M., Ardeshiri, T., Kontoyiannis, I., and Godsill, S. J. (2018b). Sharp Gaussian Approximation Bounds for Linear Systems with α-stable Noise. In *2018 IEEE International Symposium on Information Theory Proceedings*. 118, 131, 141, 148

Riabiz, M. and Godsill, S. J. (2017). Approximate simulation of linear continuous time models driven by asymmetric stable Lévy processes. In *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*. 163

Riabiz, M., Lindsten, F., and Godsill, S. (2015). Pseudo-Marginal MCMC for Parameter Estimation in α-Stable Distributions. *IFAC-PapersOnLine*, 48(28):472 – 477. 17th IFAC Symposium on System Identification SYSID 2015. 84, 95

Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Springer. 42, 47, 48, 52, 60, 91

Robert, C. P. (1995). Simulation of truncated normal variables. *Statistics and Computing*, 5(2):121–125. 94

Roberts, G. O., Rosenthal, J. S., et al. (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Statistical science*, 16(4):351–367. 49

Rosiński, J. (2001). Series representations of Lévy processes from the perspective of point processes. In Barndorff-Nielsen, O., Resnick, S., and Mikosch, T., editors, *Lévy Processes*, pages 401–415. Birkhauser Boston. 25

Roy, T. (2012). The Stable Limit Theorem. Technical report, University of Ottawa. 5

Royston, J. P. (1983). Some techniques for assessing multivarate normality based on the shapiro- wilk w. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 32(2):121–133. 182

Rudin, W. (1976). *Principles of mathematical analysis*. McGraw-hill New York. 198

Rudin, W. (2006). *Real and complex analysis*. Tata McGraw-Hill Education. 110

Samoradnitsky, G. and Taqqu, S. (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Stochastic Modeling Series. Taylor & Francis. 13, 14, 15, 16, 18, 19, 20, 21, 23, 27, 28, 29, 32, 33, 34, 35, 163, 164

Sato, K. (1999). *Lévy processes and infinitely divisible distributions*. Cambridge university press. 23

Schön, T., Gustafsson, F., and Nordlund, P. J. (2005). Marginalized particle filters for mixed linear/nonlinear state-space models. *IEEE Transactions on Signal Processing*, 53(7):2279–2289. 10

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611. 182

Shevlyakov, G. and Kim, K. (2006). Robust minimax detection of a weak signal in noise with a bounded variance and density value at the center of symmetry. *IEEE Transactions on Information Theory*, 52(3):1206–1211. 3

Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika*, 42(3/4):425–440. 3

Sornette, D. (2006). *Critical phenomena in natural sciences: chaos, fractals, selforganization and disorder: concepts and tools*. Springer Science & Business Media. 3

Stoev, S. A., Michailidis, G., and Taqqu, M. S. (2011). Estimating heavy-tail exponents through max self-similarity. *IEEE Transactions on Information Theory*, 57(3):1615–1636. 8

Stumpf, M. P. H. and Porter, M. A. (2012). Critical truths about power laws. *Science*, 335(6069):665–666. 3

Tankov, P. and Cont, R. (2015). *Financial Modelling with Jump Processes, Second Edition*. Chapman and Hall/CRC Financial Mathematics Series. Taylor & Francis. 23, 24, 38

Tanner, M. A. (1997). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. Springer Series in Statistics. Springer New York. 60

Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393):82–86. 90

Tsionas, E. G. (1999). Monte Carlo inference in econometric models with symmetric Stable disturbances . *Journal of Econometrics*, 88(2):365 – 401. 9, 63, 65

Tzagkarakis, G. (2009). *Bayesian compressed sensing using α-stable distributions*. PhD thesis, Department of Computer Science, University of Crete, Crete, Greece. 3

Uchaikin, V. and Zolotarev, V. (1999). *Chance and Stability: Stable Distributions and their Applications*. Modern Probability and Statistics. De Gruyter. 6

Uhlenbeck, G. E. and Ornstein, L. S. (1930). On the Theory of the Brownian Motion. *Phys. Rev.*, 36:823–841. 26

Unser, M., amd A. Amini, P. D. T., and Kirshner, H. (2014a). A Unified Formulation of Gaussian Versus Sparse Stochastic Processes - 2014; Part II: Discrete-Domain Theory. *IEEE Transactions on Information Theory*, 60(5):3036–3051. 26

Unser, M. and Tafti, P. D. (2014). *An Introduction to Sparse Stochastic Processes*. Cambridge University Press. 3

Unser, M., Tafti, P. D., and Sun, Q. (2014b). A Unified Formulation of Gaussian Versus Sparse Stochastic Processes - 2014; Part I: Continuous-Domain Theory. *IEEE Transactions on Information Theory*, 60(3):1945–1962. 26

Van Loan, C. (1978). Computing integrals involving the matrix exponential. *IEEE transactions on automatic control*, 23(3):395–404. 175, 176

Wahlström, N., Axelsson, P., and Gustafsson, F. (2014). Discretizing stochastic dynamical systems using Lyapunov equations. *IFAC Proceedings Volumes*, 47(3):3726 – 3731. 19th IFAC World Congress. 176

Wakefield, J., Gelfand, A., and Smith, A. (1991). Efficient generation of random variates via the ratio-of-uniforms method. *Statistics and Computing*, 1(2):129–133. 66

Warren, D. J. and Thomas, J. B. (1991). Asymptotically robust detection and estimation for very heavy-tailed noise. *IEEE Transactions on Information Theory*, 37(3):475–481. 3

Wei, G. C. and Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American statistical Association*, 85(411):699–704. 60

Weron, R. (1996a). Correction to: 'On the Chambers-Mallows-Stuck method for simulating skewed Stable random variables'. MPRA Paper 20761, University Library of Munich, Germany. 16, 17

Weron, R. (1996b). On the Chambers-Mallows-Stuck method for simulating skewed Stable random variables. *Statistics & Probability Letters*, 28(2):165 – 171. 17

Winkelbauer, A. (2012). Moments and Absolute Moments of the Normal Distribution. *ArXiv e-prints*. 190, 194

Zolotarev, V. M. (1966). On representation of Stable laws by integrals. *Selected Translations in Mathematical Statistics and Probability*, 6(1):84–5. 30

Zolotarev, V. M. (1986). *One-Dimensional Stable Distributions*. Translations of mathematical monographs. American Mathematical Society. 14, 16

Zolotarev, V. M. (1995). On representation of densities of stable laws by special functions. *Theory of Probability & Its Applications*, 39(2):354–362. 6